

The Impact of Generative AI Tools on Academic Performance in Higher Education: A Systematic Review (2023–2025)

P. Rajesh Kumar, Safiah Md. Yusof, Raziana Che Aziz, Alias Mahmud

Open University Malaysia (OUM), Malaysia

Email: rajesh@oum.edu.my, safiah_md.yusof@oum.edu.my, raziana@oum.edu.my, aliasmahmud@oum.edu.my

DOI Link: <http://dx.doi.org/10.6007/IJARBSS/v15-i12/27076>

Published Date: 22 December 2025

Abstract

This paper examines how generative AI tools—such as ChatGPT, GitHub Copilot, and related large language models—are influencing academic performance in higher education. We reviewed empirical studies published between 2023 and 2025 and followed PRISMA guidelines to search major open-access databases, including ERIC, DOAJ, arXiv, MDPI, and Semantic Scholar. Our review focused on research that reported measurable outcomes such as GPA, exam scores, writing quality, and coding accuracy. In total, 14 studies met the inclusion criteria, representing a range of countries, disciplines, and instructional contexts. The results are far from uniform. Several studies reported gains in writing, coding, and task efficiency when generative AI was used in a structured and purposeful way. Others, however, found little improvement or even lower performance when students relied heavily on AI without guidance. Taken together, the evidence shows that AI's impact on learning is highly varied and strongly dependent on context. This underscores the need for deliberate integration and for more sustained, data-driven research as these tools continue to evolve.

Keywords: Generative AI tools, Academic Performance, Higher Education

Introduction

Open AI's public release of ChatGPT in 2022 was perhaps the most prominent event in artificial intelligence that year. Since then, generative AI (GenAI) tools have had a major impact in higher education. Unlike most educational technologies, GenAI tools rapidly evolve and, therefore, are accessible to a vast number of learners through the internet. Consequently, there is a need for university administrators, programme designers, and policymakers to rethink the delivery of curriculum, assessment, and academic support in this new environment. It is now possible through tools such as ChatGPT, Claude, and GitHub Copilot for users to draft essays, solve intricate problems, and even produce academic

content within minutes. The promptness of the change is not only affecting curriculum design but is also changing the way in which students interact with learning materials (Chan & Hu, 2023).

With these breakthrough technologies, students are no longer merely using GenAI tools to write their assignments. Quite a few of them use such tools for better understanding of the subject matter, exploring examples, and obtaining immediate feedback on the drafts they have written. A great number of them claim that they have become more confident and efficient especially in writing and revision tasks, and that they have also received more help than before (Chan & Hu, 2023). However, this argument is not supported by all the available evidence. Several positions take the stance that an enormous reliance on GenAI may lead to a decrease of active learning. Instead of struggling with complex ideas, students may thus become over-reliant on the AI to carry out the cognitive tasks for them. Abbas et al. (2024), for instance, in their study found that frequent users of ChatGPT tended to procrastinate, had weaker memory retention, and obtained lower grades, particularly when they were under time pressure.

Along with the growing usage of GenAI, worries about academic integrity have also been increased. Because these instruments are capable of generating essays or computer codes in almost a split second, the authenticity of the traditional take-home assessments could be challenged. Cotton et al. (2024) maintain that, on the whole, academic integrity is facing a crisis, whereas Kofinas et al. (2024) suggestion is that assessment strategies should be reconsidered instead of AI use being entirely banned. Similarly, organizations responsible for regulation like TEQSA (2024) have suggested that such changes in the educational landscape should be reflected in the timely modification of institution-level policies.

Students' quick adaptation to GenAI technologies is one of the factors that intensify these problems. The HEPI Student AI Survey (2025) reported that, as a result of students' initiative, GenAI had already become an integral part of their coursework in many cases and that most of the learners had received no formal training and lacked instruction from teaching staff. The reaction of institutions is very different, some universities, for example, have enforced very rigorous constraints on the use of AI in some cases even before determining its actual impact on learning through sufficient research (Wang et al., 2024).

Whereas opinion pieces and policy discussions regarding GenAI are rapidly increasing, there is still little empirical evidence about the impact of GenAI on student performance. Most of the early debates were dominated by speculations and commentaries instead of research data. Only lately have more thorough, evidence-based studies come to light. Some of these studies point to improvements in the quality of writing and learning efficiency when GenAI is used intentionally (Lee, 2024). On the other hand, some raise concerns like the decline of critical thinking skills and the misuse of AI-generated content. What is missing is a clear integration of these new studies.

This research paper intends to fill that void. It is a review of empirical, open-access studies published between 2023 and 2025 that investigate the effect of GenAI on academic performance in higher education. The mission of the project is to equip educators and decision-makers with a set of insights that are not only grounded in evidence but also capable

of informing assessment design, teaching practice, and policy development as universities move forward in the digital age of learning.

Review Design

This study adopted a systematic literature review (SLR) approach guided by the PRISMA 2020 framework (Page et al., 2021). The goal was to consolidate recent empirical evidence on the impact of generative AI tools on academic performance among students in higher education. Following established SLR protocols (Joanna Briggs Institute, 2020), the process included four key phases: identification, screening, eligibility assessment, and inclusion.

Given the conceptual diversity of GenAI research, a narrative synthesis was employed rather than a meta-analysis, allowing integration of findings across varied designs and outcome measures (Grant & Booth, 2009). The review focused exclusively on open-access studies to ensure transparency and replicability.

Eligibility Criteria

Studies were included if they met the following criteria:

Inclusion Criteria

- Published between 2023 and 2025
- Open-access and available in full text
- Written in English
- Focused on higher-education learners
- Empirical and quantitative in design
- Examined a generative AI tool (e.g., ChatGPT, Bard, Claude, GitHub Copilot)
- Reported at least one measurable academic-performance outcome, such as
- GPA
- exam/test scores
- writing quality
- coding accuracy
- task efficiency
- assignment grades

Exclusion Criteria

- Qualitative-only studies
- Conceptual, theoretical, or commentary papers
- Studies focusing on K–12 or non-student populations
- Paywalled or inaccessible articles
- Papers without performance-related outcomes

Information Sources and Search Strategy

A structured search was conducted across eight open-access databases and repositories to ensure full accessibility of all included studies. The following sources were selected for their comprehensive coverage of education, linguistics, computing, and interdisciplinary scholarship:

- ERIC (Education Resources Information Center)
- DOAJ (Directory of Open Access Journals)
- CORE.ac.uk

- Semantic Scholar (filtered for open-access articles only)
- SpringerOpen
- Frontiers (fully open-access platform)
- MDPI
- PLOS ONE
- arXiv (preprints relevant to computing and education)

Searches covered the period 2023–2025, reflecting the rapid rise of generative-AI tools in higher education. Boolean combinations were used to maximise retrieval of relevant studies, for example:

("ChatGPT" OR "generative AI" OR "large language model")
 AND ("academic performance" OR "learning outcomes" OR "assessment")
 AND ("higher education" OR "university")
 ("ChatGPT" OR "generative AI" OR "large language model")
 AND ("academic performance")
 AND ("higher education" OR "university")

Only empirical studies with open-access full text were eligible for screening.

Although Elsevier, Wiley, Taylor & Francis, and Sage were *not* used as search databases, open-access versions of eligible studies published by commercial journals were included when publicly available under Creative Commons licences. This approach ensured that the review remained consistent with its commitment to open-access research while capturing all relevant, legally accessible empirical evidence.

Screening and Selection

Our initial search across specified databases yielded a total of 85 records. After removing 12 duplicate records, 73 unique studies remained for title and abstract screening. At this stage, 36 studies were excluded for the following reasons: 17 studies were qualitative-only or conceptual commentaries. 8 studies did not focus on higher education populations. 11 studies lacked any objective, measurable academic performance outcomes.

This left 37 studies for which we sought to full text to assess for eligibility. After a detailed review, 23 studies were excluded at this final stage for specific reasons: 9 studies had mixed-methods or qualitative-dominant designs that did not allow for quantitative synthesis of performance outcomes. 4 studies were conceptual or theoretical pieces despite initial classification. 4 studies focused on non-student populations. 2 studies had an inaccessible or incomplete full text. 6 studies, while quantitative, did not report on a specific academic performance metric. Ultimately 14 studies met all inclusion criteria and were retained for the final synthesis.

Table 1

PRISMA Overview

Records identified	85
After duplicates removed	73
Titles/abstracts screened	73
Excluded	36
Full-text articles assessed	37
Studies included	14

Data Extraction

For each of the 14 included studies, the following details were extracted using a standardized form: author(s), year, country, database/journal, discipline, AI tool investigated, study design, sample size, type of performance measure, and direction of effect. A summary of these extracted variables is presented in Table 1.

Table 1

Data Extraction and Synthesis Summary of Verified Open-Access Studies (n = 14)

Author(s) & Year	Country	Database / Journal	Discipline	AI Tool(s)	Design / Sample (n)	Performance Outcome Measured	Effect
Wang & Fan, 2025	China	Hum. & Soc. Sci. Communications	Education	ChatGPT	Meta-analysis (51 studies)	Learning performance (various measures)	Large positive impact on learning performance (Hedges' $g = 0.867$).
Fan et al., 2025	China	Scientific Reports (DOAJ)	Engineering Education	Generative AI	Survey of students (n=148)	Self- reported learning efficiency, initiative, creativity; academic performance (grades)	>50% of students reported positive impact on learning efficiency, initiative and creativity (and ~50% on independent thinking), but no change in actual academic performance.
Vieriu et al., 2025	Romania	Education Sciences (MDPI)	Educational Science/Eng.	AI (incl. ChatGPT)	Survey of students (n=85)	Perceived impact on grades (exams, projects)	82.4% of respondents believed AI use enhanced their academic performance (exams/projects); 15.3% saw no change.

Author(s) & Year	Countr y	Database / Journal	Discipline	AI Tool(s)	Design / Sample (n)	Performanc e Outcome Measured	Effect
Wecks et al., 2024	Germa ny	arXiv preprint	Economics/Educatio n	Generati ve AI	Essay analysis (cohort size NR)	Exam performanc e (score)	Students identified as GenAI users scored on average 6.71 points lower (out of 100) than non- users, suggesting generative AI use correlated with poorer exam outcomes. Students with moderate chatbot use achieved better exam outcomes; excessive use yielded no additional gain.
Sánchez- Vera et al., 2024	Spain	Education Sciences (MDPI)	Education (Early Childhood)	LLM- based Chatbot	Pilot study (n=42 students)	Exam scores (final exam)	Frequent ChatGPT use was negatively related to CGPA ($\beta = -$ 0.104, $p < 0.05$), i.e. higher use corresponded to lower academic performance.
Abbas et al., 2024	Pakista n	Int. J. Educ. Technol. in Higher Ed.	Management/Educa tion	ChatGPT	Time-lagged survey Study 1(n=165) Study 2(n=494)	Academic performanc e (CGPA)	Positive (more ChatGPT use linked to higher grades) No effect (no significant grade difference between ChatGPT users and non- users)
López Fernández & Vergaz (2024)	Spain	AI (MDPI)	Computer Science	ChatGPT	Exploratory correlational study – N=37 CS students	Course grades (database admin)	Positive (students’ perceptions of
Mennella & Quadros- Mennella (2024)	USA	J. Univ. Teach. Learn. Pract. (JUTLP, OER)	Education (Writing Instruction)	ChatGPT	Survey/analysi s – N=49 (undergraduat es)	Assignment grades	
Mehrvarz et al. (2025)	Iran	Computers & Educ.: Artificial	Educational Technology	ChatGPT	Survey – N=305	Self- reported academic	

Author(s) & Year	Country	Database / Journal	Discipline	AI Tool(s)	Design / Sample (n)	Performance Outcome Measured	Effect
		Intelligence (Elsevier)			university students	performance	ChatGPT positively influenced performance)
Shihab et al. (2025)	Canada	arXiv Preprint	Computer Science (Programming)	GitHub Copilot	Experimental / 156	Coding accuracy, time efficiency	↑
Sultan et al. (2025)	Pakistan	Journal of Arts and Linguistics Studies (DOAJ)	EFL / Applied Linguistics	ChatGPT	Quasi-experimental / 78	Essay structure and coherence	↑
Yusof (2025)	Malaysia	J. Inf. Technol. Educ.: Research (OER)	Education (Measurement & Eval.)	ChatGPT	Retrospective cohort – 2nd-year students; N=64 control, 64 experimental	Final exam scores	Positive (ChatGPT-assisted group scored significantly higher)
Khampusa et al. (2025)	Thailand	LEARN J. (Lang. Educ. & Acquisition)	Language Education (EFL)	ChatGPT	Mixed-methods – N=30 (third-year English majors)	Writing quality (argumentative essays)	Positive (significant improvements in writing performance)
Youssef et al. (2025)	UAE	Comput. & Educ.: Artificial Intelligence	Education (General)	ChatGPT	Survey – N=254	GPA/Self-reported learning)	Negative correlation between frequency of ChatGPT usage and GPA scores

Synthesis

Given such a diverse range of designs from experiments, surveys, and meta-analyses, a narrative synthesis was used to bring together the key findings. Educational outcomes were very different from one study to another, including such things as test scores, GPA, quality of writing, and correctness of programming. From this variation, we have identified three general themes:

Overall Impact on Academic Performance

When we compare such research as those side by side, it is difficult to come up with a simple answer as to whether generative AI is good or bad. The evidence is quite mixed, actually. On the bright side, the meta-analysis by Wang and Fan (2025) indicates that AI-based tools such as ChatGPT tend to have an average positive effect on learning performance across different studies. The same idea is brought forward by Mehrvarz et al. (2025) where students having a positive attitude towards ChatGPT are more engaged and thus have better overall achievements. Besides, there are positive results from the highly targeted experiments as well: Sánchez-Vera (2024) demonstrated how a subject-area chatbot could elevate students'

scores, and Yusof (2025) shared data on higher test scores when ChatGPT was used as a tool for retrieval-based learning activities.

However, the method is not free from problems. Abbas et al. (2024), for instance, discovered that those who frequently used ChatGPT had notably lower GPAs, particularly students who were already applying in a negative way procrastination or feeling time pressure. The same kind of situation, i.e., generative AI, causing to drop in exam scores, was observed by Wecks et al. (2024). Students who used generative AI as a part of their test preparation scored almost seven points less than those who didn't, the average difference. On a different note, Mennella and Quadros-Mennella (2024) argue that in a few cases - especially in writing - the involvement or non-involvement of AI tools even makes no difference to the obtained grades. All of these pieces of evidence indicate that generative AI is far from being a magic wand. Its impact on academic success seems to hinge greatly on how students use it and under what kind of condition is present.

Discipline-Specific Patterns

Besides, there are very significant changes in the behaviour of different disciplines. The pattern is not consistent. In technical fields, particularly in programming, the results are mainly positive. For example, Shihab et al. (2025) revealed that GitHub Copilot was instrumental in helping students to write accurate as well as efficient code. By the same token, López Fernández and Vergaz (2024) pointed out that the employment of ChatGPT led to better performance in courses related to database administration.

On the other hand, the humanities-oriented students also can benefit from the well-structured use of AI. Sultan et al. (2025) demonstrated that the usage of ChatGPT as a component of the writing process led to the improvement of the structure as well as the coherence of the essays of undergraduate English students. Khampusaen (2025) also similarly confirmed significant improvement in the quality of persuasive writing by English language learners. These complete, discipline-targeted samples accompany general studies that show neutral as well as negative effects. This opposition is helpful in recognizing the importance of the context and the task. When AI is associated with a definite goal and clear learning objectives it generally facilitates learning; when students use it in an unraveled, improvised manner, then they cannot be sure that the benefits will be there.

Factors that Influence Effectiveness

One of the themes that appear repeatedly in this research is the way students' opinion and usage of generative AI affects educational benefits considerably. Mehrvarz et al. (2025) report that those students who consider ChatGPT as a helpful and reliable tool are found to have better academic achievements overall. On the flip side, situations with negative or weaker outcomes usually imply that AI use is less structured. Abbas et al. (2024) and Wecks et al. (2024) are examples of such cases, where students were left on their own to figure out how to use AI without much support or guidance from teachers.

In the research of Sánchez-Vera (2024), the chatbot was part of a well-arranged educational system rather than being given as an optional extra. Yusof (2025) implemented ChatGPT to promote specific active learning methods, such as retrieval practice, instead of just a general-purpose assistant. Vieriu and Petrea (2025) also point out that unstructured, heavy

consumption of a digital resource may lead to over-reliance, decrease of critical thinking skills, and increase of academic dishonesty. Collectively, these pieces of research lead to one main conclusion: the pedagogical worth of generative AI is less dependent on the tool itself and more on how intentionally it is merged with teaching and learning.

Discussion

We argued in this review that generative AI does not influence academic performance in the same way, and its effect varies depending on usage. AI can be a great help to students when it is integrated into well-structured, discipline-specific tasks, such as coding (Shihab et al., 2025), retrieval practice (Yusof, 2025), or targeted writing support (Sultan et al., 2025). Here, AI is a tool that helps the students develop their own thinking, which is not replaced by the machine.

On the other hand, if AI is used in an unregulated manner, thus for instance, as a general saving of time for homework, the consequence will be a drop in the output quality or, at best, a stagnation of the performance level of the user. Abbas et al. (2024) and Weeks et al. (2024) studies indicate that along with cognitive “offloading” and disengagement from the depth of the task, performance is lowered. The evidence cited by the authors corroborates the warning raised by Vieriu and Petrea (2025) about the risk of overdependence on AI that leads to a decrease in critical thinking. The authors of the paper are quite in line with the idea that moral dilemmas presented by AI should not be reduced to a binary marking the boundary between “good” and “bad”, but rather determining under what circumstances technology leads to better or worse learning. The efficacy of the technology is mostly determined by whether the use of AI is adequately prepared for the learning activities, which includes facilitating the learning by the teacher as well as AI literacy skills that pay attention to the users as critical and ethical individuals.

Our results have to be interpreted considering several limitations. The first consideration goes to the methodology of the studies included in the review that varied widely, including surveys and quasi-experiments to meta-analyses, and also had different outcome measures. Differences between these works made it impossible to establish clear cause effects or calculate the effect sizes consistently, therefore we used narrative synthesis which was shaped more by the different kinds of contrast e.g., structured vs. self-directed use. The second limitation is related to the fact that the review was limited to sources that are fully open-access, so the evidence comes primarily from the regions and establishments which publish in OA journals and are often early adopters of educational AI. This may lead to an overestimation of the number of radical, course-embedded intervention (e.g., coding classes, discipline-specific chatbots) and a low representation of more traditional or less resourced educational institutions.

Another limitation is the issue of time since almost all the studies were done in the time span from 2023 to 2025, a time of rapid change in AI capability, student familiarity, and institutional policy. Most of the negative effects reported are due to the first, unguided uses of AI, and this trend may reverse as AI literacy increases and more structured integration becomes the norm. Lastly, studies varied in their manner of measuring performance; some used GPA, which was self-reported, while others used exam scores, or code quality, thus making it hard to compare them directly. These limitations signal to our main point: the potential of

generative AI in education mostly depends on the context. The patterns we point out should be seen as taking effect; they will be reassessed when both AI tools and teaching practices keep developing.

Conclusion

In this work, we have attempted to understand a rapidly changing and, at times, inconsistent research area. Our main discovery is that generative tools can be authentic enablers for student learning provided that we integrate them into well-defined, scaffolded learning activities—coding exercises, retrieval practice, or guided writing tasks—where they function as thinking helpers rather than substitutes. On the contrary, if students employ them in a lax manner, mostly with the aim of "getting the work done," the situation turns out to be completely different: performance is lowered most frequently, and deep engagement is weakened.

The main message from this study is simple: what matters is not the Gen AI tool on its own, but how we choose to use it. There are three factors behind the successful use of the Gen AI tools that we can actually plan for: the alignment of the tool and the task, providing students with guidance on the device utilization, and teaching them to evaluate and use it in an ethical manner. Our work is subject to limitations—only open-access studies, different methods, and a short time window during which the technology was changing—but it, nevertheless, suggests a clear direction. If we aim to have Gen AI tools be a source of better performance instead of a reason for its decline, then we must treat them as one of the teaching methods rather than a way around pedagogy.

Contributions of the Present Study

This study contributes to the ongoing conversation about generative AI in higher education in two key ways. Theoretically, it helps clarify why the impact of tools like ChatGPT appears so uneven across the literature. Instead of treating AI as a single, uniform intervention, the review shows that outcomes depend heavily on how these tools are woven into the learning process—whether they support structured tasks, promote active engagement, or, conversely, encourage shortcuts and cognitive offloading. By bringing these patterns together, the study offers a clearer explanation for the contradictory findings reported in earlier research and adds a more grounded understanding of how AI interacts with student learning. Contextually, the review pulls together evidence from different countries, disciplines, and institutional settings at a moment when universities are still working out their policies and practices. This broader perspective highlights the importance of AI literacy, assessment design, and institutional guidance—factors that often get overlooked in more tool-focused discussions. Taken together, these contributions strengthen the evidence base for educators and policymakers and provide a more realistic picture of where generative AI can genuinely support learning and where it may introduce new challenges.

References

- Abbas, M., Jam, F. A., & Khan, T. I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21(1), 10. <https://doi.org/10.1186/s41239-024-00444-7>
- Ashraf, M. A., Alam, J., & Kalim, U. (2025). Effects of ChatGPT on students' academic performance in Pakistan higher education classrooms. *Scientific Reports*, 15(1), 16434. <https://doi.org/10.1038/s41598-025-92625-1>
- Alammari, D., Alamari, E., Alamri, R., Alharbi, R., Felimban, J., & Aljohani, J. (2025). Assessing medical students' attitudes, performance, and usage of ChatGPT in Jeddah, Saudi Arabia. *Frontiers in Artificial Intelligence*, 8, 1577911. <https://doi.org/10.3389/frai.2025.1577911>
- Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. *International Journal of Educational Technology in Higher Education*, 20, 43. <https://doi.org/10.1186/s41239-023-00411-8>
- Chen, J., Zhuo, Z., & Lin, J. (2023). Does ChatGPT play a double-edged sword role in the field of higher education? An in-depth exploration of the factors affecting student performance. *Sustainability*, 15(24), 16928. <https://doi.org/10.3390/su152416928>
- Cotton, D. R. E., Cotton, P., & Shipway, R. (2024). Chatting and cheating: Ensuring academic integrity in the age of ChatGPT. *Innovations in Education and Teaching International*, 61(1), 1–12. <https://doi.org/10.1080/14703297.2023.2190148>
- Fan, L., Deng, K., & Liu, F. (2025). Educational impacts of generative artificial intelligence on learning and performance of engineering students in China. *Scientific Reports*, 15, 26521. <https://doi.org/10.1038/s41598-025-06930-w>
- Freeman, J. (2025). *Student generative AI survey 2025*. Higher Education Policy Institute.
- Khampusaen, D. (2025). The impact of ChatGPT on academic writing skills and learning outcomes in EFL argumentative essays. *LEARN Journal*, 18(1), 963–988. <https://files.eric.ed.gov/fulltext/EJ1470999.pdf>
- Kofinas, A. K., Tsay, C.-H.-H., & Pike, D. (2025). The impact of generative AI on academic integrity of authentic assessments within a higher education context. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.13585>
- Lee, D., Arnold, M., Srivastava, A., et al. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers & Education: Artificial Intelligence*, 6(2), 100221. <https://doi.org/10.1016/j.caeai.2024.100221>
- Lodge, J. M. (2024a). The evolving risk to academic integrity posed by generative artificial intelligence. Tertiary Education Quality and Standards Agency. <https://www.teqsa.gov.au/sites/default/files/2024-08/evolving-risk-to-academic-integrity-posed-by-generative-artificial-intelligence.pdf>
- López Fernández, D., & Vergaz, R. (2024). Adoption and impact of ChatGPT in computer science education: A case study on a database administrative course. *arXiv*. <https://arxiv.org/abs/2407.12145>
- Mehrvarz, M., Salimi, G., Abdoli, S., & McLaren, B. M. (2025). How does students' perception of ChatGPT shape online learning engagement and performance? *Computers & Education: Artificial Intelligence*, 100459. <https://doi.org/10.1016/j.caeai.2025.100459>

- Mennella, T., & Quadros-Mennella, P. (2024). Student use, performance and perceptions of ChatGPT on college writing assignments. *Journal of University Teaching and Learning Practice*, 21(1), 95–118.
- Page, M. J. (2021). The PRISMA 2020 statement. *BMJ*, 372, n71 <https://doi.org/10.1136/bmj.n71>
- Ravšelj, D., Keržič, D., Tomaževič, N., Umek, L., Brezovar, N., Iahad, N. A., et al. (2025). Higher education students' perceptions of ChatGPT: A global study of early reactions. *PLOS ONE*, 20(2), e0315011. <https://doi.org/10.1371/journal.pone.0315011>
- Sánchez-Vera, F. (2024). Subject-specialized chatbot in higher education as a tutor for autonomous exam preparation: Analysis of the impact on academic performance and students' perception of its usefulness. *Education Sciences*, 15(1), 26. <https://doi.org/10.3390/educsci15010026>
- Shihab, A., Wang, J., & Fard, A. (2025). The effects of GitHub Copilot on computing students' programming effectiveness, and processes in brownfield programming tasks. *arXiv Preprint*. <https://arxiv.org/abs/2506.10051>
- Shihab, A. (2025). Perceptions of generative AI tools in higher education: Insights from students and academics at Sultan Qaboos University. *Education Sciences*, 15(11), 1464. <https://doi.org/10.3390/educsci15111464>
- Sultan, M., Taj, A., Sabir, F., Ali, S., & Qureshi, M. O. K. (2025). The effect of ChatGPT on the writing abilities of undergraduate English students. *Journal of Arts and Linguistics Studies*, 2(4), 67–91. <https://doi.org/10.71281/jals.v3i1.210>
- Sun, D., Boudouaia, A., Zhu, C. (2024). Would ChatGPT-facilitated programming mode impact college students' programming behaviors, performances, and perceptions? An empirical study. *International Journal of Educational Technology in Higher Education*, 21, 14. <https://doi.org/10.1186/s41239-024-00446-5>
- Vieriu, A. M., & Petrea, G. (2025). The impact of artificial intelligence (AI) on students' academic development. *Education Sciences*, 15(3), 343. <https://doi.org/10.3390/educsci15030343>
- Wang, H., Dang, A., Wu, Z., & Mac, S. (2024). Generative AI in higher education: Seeing ChatGPT through universities' policies, resources, and guidelines. *arXiv*. <https://arxiv.org/abs/2312.05235>
- Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications*, 12(1), 412. <https://doi.org/10.1057/s41599-025-04787-y>
- Wecks, J. O., Voshaar, J., Plate, B. J., & Zimmermann, J. (2024). Generative AI usage and exam performance. *SSRN Working Paper*. <https://doi.org/10.2139/ssrn.4812513>
- Youssef, E., Medhat, M., Abdellatif, S., & Al Malek, M. (2024). Examining the effect of ChatGPT usage on students' academic learning and achievement: A survey-based study in Ajman, UAE. *Computers & Education: Artificial Intelligence*, 7, 100316. <https://doi.org/10.1016/j.caeai.2024.100316>