HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

GENERATIVE ARTIFICIAL INTELLIGENCE AND STUDENTS' INDEPENDENT LEARNING IN PROGRAMMING EDUCATION: A NARRATIVE REVIEW OF EMPIRICAL EVIDENCE

Ilhom Jumayev

Denau Institute of Entrepreneurship and Pedagogy, Denov, Uzbekistan
Termez State Pedagogical Institute (independent researcher), Termez, Uzbekistan

Doi: https://doi.org/10.5281/zenodo.23373491

Abstract. Background: Generative artificial intelligence (AI) tools are now widely available to students who learn programming, yet it remains unclear whether higher task performance with AI leads to durable, independent learning. **Methods:** A narrative review of empirical studies published in 2023–2025 was carried out in September–October 2026. Studies were included if generative AI was used in a learning task and if performance with AI, performance without AI or learning processes were measured. Five controlled or experimental studies formed the core evidence; one qualitative study, one classroom deployment and two surveys were used as supplementary evidence. All bibliographic data were verified against publisher records. **Results:** AI consistently raised performance while it was available, but its effect on unassisted performance ranged from negative to neutral. Negative effects appeared when learners obtained complete solutions, whereas hint-based or explanation-oriented use reduced or removed them. Several studies suggest that AI may widen the gap between learners with high and low prior knowledge and weaken self-monitoring. Only one study measured delayed retention, and only one measured transfer. Programming courses should therefore assess unassisted performance separately and organise AI use in guided stages.

Keywords: AI-assisted learning; cognitive offloading; computing education; generative artificial intelligence; independent learning; informatics teaching methodology; programming education; transfer of learning.

INTRODUCTION

Large language models that generate program code have entered programming classrooms in a very short time. Such models solve many typical problems of introductory programming courses, and this has led computing educators to reconsider what students should learn and how their learning should be assessed [2]. In Uzbekistan, the development of AI technologies has been declared a strategic priority until 2030 [14], and AI-related topics are being introduced into the school informatics curriculum [17]. National studies have examined the methodological foundations of AI in informatics education, mainly from the perspective of individualised and adaptive instruction [6], and have argued for combining fundamental, professional and general-education approaches when preparing informatics teachers to work with AI [18].

A central question, however, is still open: does better performance with AI mean that students have learned more? Two outcomes have to be distinguished. The first is the quality of a product created with AI assistance. The second is what the student can do independently once the tool is removed. The concept of cognitive offloading explains why these outcomes may diverge. Offloading a

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

mental task to an external tool reduces internal cognitive demand and can make a task easier [15]. In experimental work, offloading improved immediate performance but weakened later memory of the offloaded information [5]. With generative AI, students can offload not only information but also problem analysis, algorithm selection and the justification of a solution.

This review addresses three questions. (1) How does generative AI affect performance with AI and performance without AI in programming and related learning tasks? (2) Which conditions of use moderate these effects? (3) What do the findings imply for the teaching of informatics? The review is part of the author's doctoral research on developing students' intellectual-creative abilities with the help of AI [7].

MATERIALS AND METHODS

Design. The study is a narrative review. It was carried out in September–October 2026 at the Denau Institute of Entrepreneurship and Pedagogy (Uzbekistan) as part of doctoral research conducted at Termez State Pedagogical Institute.

Search and selection. Studies were identified in Google Scholar, the ACM Digital Library, publishers' websites and arXiv using combinations of the terms "generative AI", "ChatGPT", "code generator", "novice programmers", "learning outcomes" and "critical thinking". Studies were included if they (a) were published in 2023–2025; (b) used generative AI in a learning or problem-solving task; and (c) measured at least one of the following: performance with AI, performance without AI, or learning process data. Studies with a control or comparison condition and an unassisted outcome were classified as core evidence. Qualitative, deployment and survey studies were classified as supplementary evidence.

Coding and verification. For each study, the participants and context, the design, and the presence of five types of measures were recorded: performance with AI, unassisted performance, delayed retention, transfer, and process or self-report data. Bibliographic details were checked against DOI records and publisher pages. One study is a preprint that has not been peer-reviewed; it is marked as such. The review is not systematic, the number of included studies is small, and the coding was done by one reviewer; the conclusions are therefore stated with caution.

RESULTS AND DISCUSSION

Five studies met the criteria for core evidence and four studies were used as supplementary evidence. Table 1 summarises the core studies, and Figure 1 shows which outcomes each included study measured.

Table 1

Core empirical studies: performance with and without generative AI

Study	Participants and context	Design	Performance with AI	Unassisted performance

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Study	Participants and context	Design	Performance with AI	Unassisted performance
Bastani et al. [1]	About 1,000 students, grades 9–11, one school in Turkey; mathematics	Field experiment randomised by class; three arms	Practice scores: +48% (plain GPT-4), +127% (hint-based tutor) vs. control	Closed-book exam: −17% for plain GPT-4; no significant difference for tutor
Kazemitabaar et al. [8]	69 novices aged 10–17; introductory Python	Controlled study; ten sessions	Code correctness: 80.1% vs. 44.4%	No significant difference; higher-pretest AI users retained more after one week
Xue et al. [19]	48 students in an introductory Java course (CS1), USA	Controlled two-group study	No significant difference	No significant difference (closed-book)
Fan et al. [3]	117 university students; essay writing in English	Randomised lab experiment; four groups	Largest essay-score gain in the AI group	No difference in knowledge gain or transfer
Lehmann et al. [12] (preprint)	University students learning Python; the Netherlands and Germany	Field panel and two randomised lab experiments	More topics covered when solutions were requested	No significant overall difference; solution-seeking lowered and explanation-seeking improved understanding

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

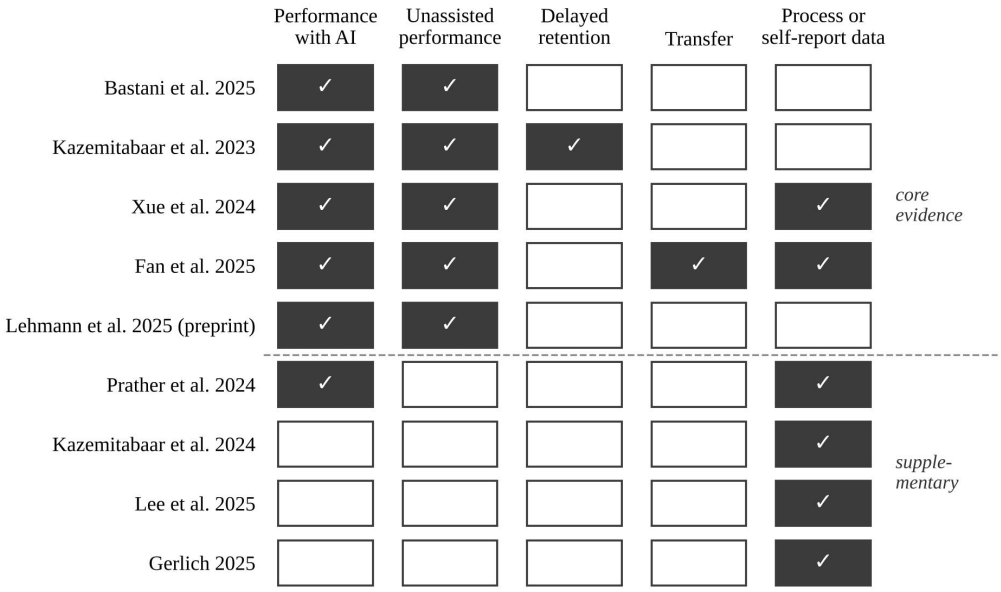


Figure 1. Outcomes measured in the reviewed studies (filled cell = measured)

Performance with AI versus independent performance

The most consistent finding is that generative AI improves performance while it is available. In a field experiment with about 1,000 high-school students, practice scores rose by 48% with a plain GPT-4 interface and by 127% with a tutor version that gave hints instead of answers [1]. Novices who used a code generator wrote correct code far more often than the control group (80.1% vs. 44.4%) [8]. In a writing task, the AI group improved its essays the most [3].

Effects on unassisted performance were weaker and less consistent. In the high-school experiment, students who had practised with the plain GPT-4 interface scored 17% lower than the control group on a closed-book exam [1]. In the programming studies, differences in unassisted test scores were not significant [8; 19], and the AI group in the writing study did not gain more knowledge or transfer it better [3]. The lab experiments in the preprint by Lehmann et al. [12] also found no significant overall effect on learning. Taken together, the evidence ranges from negative to neutral: gains with AI did not automatically become gains without AI.

The mode of AI use as a moderator

The way AI is used appears to matter more than its mere availability. In the high-school experiment, the hint-based tutor avoided the drop in exam scores observed with the plain interface [1]. In the preprint by Lehmann et al. [12], asking AI for complete solutions increased the number of topics covered but reduced understanding, whereas asking for explanations improved it. A classroom deployment of an AI programming assistant with more than 700 students led its designers to recommend avoiding direct answers, keeping students cognitively engaged and keeping the assistant transparent and under student control [9]. These results suggest that the design of AI assistance is a pedagogical decision, not only a technical one.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Prior knowledge and the widening gap

Several studies indicate that AI does not benefit all learners equally. In a qualitative study of 21 introductory programming students, about half used AI to speed up writing code they had already planned, while the other half struggled; for the latter group, AI compounded existing metacognitive difficulties and created an illusion of competence [13]. Novices with higher pre-test scores who had used a code generator performed better on the one-week retention test [8], and the preprint by Lehmann et al. [12] reported that AI use widened the gap between students with low and high prior knowledge. Because the samples are small, this pattern should be treated as a hypothesis for further testing rather than an established effect.

Self-regulation and critical thinking

Evidence on self-regulation and critical thinking is suggestive but weak. Fan et al. [3] observed differences in self-regulated learning processes between groups and interpreted them as possible "metacognitive laziness"; this is an interpretation rather than a directly measured construct. In a survey of 319 knowledge workers, higher confidence in AI was associated with less reported critical thinking [11]. A survey of 666 participants in the United Kingdom found a negative correlation between AI use and self-reported critical thinking, partly mediated by cognitive offloading [4]. Both surveys are correlational and rely on self-report, so they do not show causal effects.

Theoretical interpretation

The findings are consistent with research on cognitive offloading: external help improves immediate performance but may reduce what is retained [5; 15]. Research on productive failure offers a possible remedy. A meta-analysis of 53 studies found a moderate advantage for problem solving followed by instruction over the reverse order (Hedges' $g = 0.36$), although the effect reversed for younger learners and for domain-general skills [16]. At the same time, minimally guided instruction has been criticised as ineffective for novices [10]. Combining these positions, a short and well-scoped independent attempt followed by guided support appears more defensible than either unrestricted AI use or prolonged unguided work.

Gaps in the evidence

Figure 1 shows that only one study measured delayed retention (after one week) and only one measured transfer to a new task. Most studies lasted a single session or a few weeks, and only some involved university programming courses. Long-term effects of AI on independent programming ability therefore remain largely unknown, and evidence from Central Asian higher education is not yet available in the reviewed literature.

Implications for teaching informatics

On the basis of the reviewed evidence, the following principles are proposed for programming courses. They are recommendations derived from the review and require empirical testing.

1. Assess unassisted performance separately from the quality of AI-assisted products, for example through a final task completed without AI.
2. For novices, prefer hint-based or explanation-oriented AI modes and restrict access to complete solutions.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

3. Ask students to make a short independent attempt before they consult AI, and adjust the length of this attempt to their prior knowledge.
4. Include metacognitive checkpoints, such as a reflection on what was done independently and a declaration of how AI was used.
5. Monitor outcomes separately for students with low and high prior knowledge so that AI use does not widen existing gaps.

The author's ongoing doctoral study is testing a staged approach based on these principles in a quasi-experiment with informatics students at three pedagogical higher education institutions in Uzbekistan; its results will be reported separately.

CONCLUSION

The reviewed studies show that generative AI reliably raises performance while it is available, but its effect on what students can do without AI ranges from negative to neutral. The mode of AI use appears to be decisive: providing complete solutions was associated with lower independent performance and understanding, whereas hint-based and explanation-oriented use reduced or removed this risk. AI may also widen differences between learners with different levels of prior knowledge. Because few studies measured retention or transfer, programming courses should evaluate unassisted performance directly and introduce AI in guided stages. Longer studies in university programming courses, including in Central Asia, are needed to test these conclusions.

References

1. Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakcı, Ö. and Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, **122**(26): e2422633122. https://doi.org/10.1073/pnas.2422633122
2. Denny, P., Prather, J., Becker, B.A., Finnie-Ansley, J., Hellas, A., Leinonen, J., Luxton-Reilly, A., Reeves, B.N., Santos, E.A. and Sarsa, S. (2024). Computing education in the era of generative AI. *Communications of the ACM*, **67**(2): 56–67. https://doi.org/10.1145/3624720
3. Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X. and Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. *British Journal of Educational Technology*, **56**(2): 489–530. https://doi.org/10.1111/bjet.13544
4. Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, **15**(1): 6. https://doi.org/10.3390/soc15010006 (Correction: *Societies*, **15**(9): 252.)
5. Grinschgl, S., Papenmeier, F. and Meyerhoff, H.S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, **74**(9): 1477–1496. https://doi.org/10.1177/17470218211008060
6. Ilyasova, Z.K. (2026). Informatika ta'limida sun'iy intellekt texnologiyalarining metodologik asoslari [Methodological foundations of artificial intelligence technologies in informatics education]. *Science and Society*, (1-3): 79–81. https://journal.ndpi.uz/ss/en/article/view/2026-1-3-027

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

7. Jumayev, I.N. (2026). Aniq va tabiiy fanlarni o‘qitishda sun’iy intellekt yordamida talabalarning intellektual-ijodiy qobiliyatlarini rivojlantirish [Developing students’ intellectual-creative abilities with artificial intelligence in teaching exact and natural sciences]. *Aniq va tabiiy fanlarni o‘qitishda sun’iy intellekt, raqamli texnologiyalar va STEAM yondashuvlar*, **1**(2): 120–123. https://ilibrary.uz/index.php/tpdk/article/view/44
8. Kazemitabaar, M., Chow, J., Ma, C.K.T., Ericson, B.J., Weintrop, D. and Grossman, T. (2023). Studying the effect of AI code generators on supporting novice learners in introductory programming. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 1–23. https://doi.org/10.1145/3544548.3580919
9. Kazemitabaar, M., Ye, R., Wang, X., Henley, A.Z., Denny, P., Craig, M. and Grossman, T. (2024). CodeAid: Evaluating a classroom deployment of an LLM-based programming assistant that balances student and educator needs. In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 1–20. https://doi.org/10.1145/3613904.3642773
10. Kirschner, P.A., Sweller, J. and Clark, R.E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, **41**(2): 75–86. https://doi.org/10.1207/s15326985ep4102_1
11. Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R. and Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. In: *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. New York: ACM. https://doi.org/10.1145/3706598.3713778
12. Lehmann, M., Cornelius, P.B. and Sting, F.J. (2025). AI meets the classroom: When do large language models harm learning? Preprint, arXiv:2409.09047v2. https://arxiv.org/abs/2409.09047
13. Prather, J., Reeves, B.N., Leinonen, J., MacNeil, S., Randrianasolo, A.S., Becker, B.A., Kimmel, B., Wright, J. and Briggs, B. (2024). The widening gap: The benefits and harms of generative AI for novice programmers. In: *Proceedings of the 2024 ACM Conference on International Computing Education Research*, Vol. 1. New York: ACM, pp. 469–486. https://doi.org/10.1145/3632620.3671116
14. President of the Republic of Uzbekistan (2024). Resolution No. PQ-358 of 14 October 2024 “On approval of the Strategy for the development of artificial intelligence technologies until 2030”. https://lex.uz/docs/-7158604
15. Risko, E.F. and Gilbert, S.J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, **20**(9): 676–688. https://doi.org/10.1016/j.tics.2016.07.002
16. Sinha, T. and Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research*, **91**(5): 761–798. https://doi.org/10.3102/00346543211019105
17. Ta’lim xabarlari (2025). Kelasi o‘quv yilidan “Informatika” fani o‘quv dasturlariga sun’iy intellektga oid mavzular kiritiladi [AI-related topics will be included in the “Informatics” curricula from the next academic year], 21 October 2025. https://talimxabarlari.uz/50841/
18. Turdaliyeva, N.A. (2025). Sun’iy intellekt elementlarini informatika ta’limiga integratsiya qilish: fundamental, kasbiy va umumta’limiy yondashuvlar [Integrating artificial intelligence elements

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

into informatics education: fundamental, professional and general-education approaches]. *NamDU ilmiy axborotnomasi*, (11). https://journal.namdu.uz/archive/2025-issue-11/maqola-66222-1759918230/pdf

19. Xue, Y., Chen, H., Bai, G.R., Tairas, R. and Huang, Y. (2024). Does ChatGPT help with introductory programming? An experiment of students using ChatGPT in CS1. In: *Proceedings of the 46th International Conference on Software Engineering: Software Engineering Education and Training*. New York: ACM, pp. 331–341. https://doi.org/10.1145/3639474.3640076