APPLICATION SCENARIOS, EFFECTIVENESS, AND RISK BOUNDARIES OF GENERATIVE AI IN CLASSROOM TEACHING

Shen Mingzhu

"Tashkent Institute of Irrigation and Agricultural Mechanization Engineers" National Research University
100000,39, Kori Niyazi street,Tashkent,Republic of Uzbekistan
https://orcid.org/0009-0002-0577-8156
E-mail:18265987700@163.com

K E Y W O R D S

generative artificial intelligence,classroom teaching, application scenarios, learning outcomes,risk boundaries, human-AI collaboration

A B S T R A C T

Generative artificial intelligence (GenAI) is rapidly penetrating classroom teaching. However, existing studies mostly focus on single scenarios or isolated issues, lacking an integrated analysis of application scenarios, actual effectiveness, and risk boundaries. This study addresses three questions: What typical application scenarios has GenAI formed in classroom teaching? What effectiveness evidence has empirical research reported? How should risk boundaries be delineated? Following the PRISMA protocol, a systematic literature analysis was conducted. A total of 2,225 records were retrieved from Web of Science, Scopus, ERIC, IEEE Xplore, and CNKI, and 35 representative studies were included after screening. A three-dimensional scenario-effectiveness-risk framework was constructed for coding and synthesis. Findings: (1) GenAI has formed stable application patterns in five scenarios: lesson preparation and resource generation, in-class teaching assistance, practice and instant feedback, personalized tutoring, and academic assessment; (2) effectiveness evidence is clearly stratified: strongest for instant feedback and personalized tutoring, mixed for writing instruction and higher-order thinking, and still exploratory for open-ended assessment, while configurational evidence indicates that effectiveness stems from human-AI collaborative instructional design rather than the technology itself; (3) five major risks (academic integrity, cognitive overreliance, hallucination, digital divide, and data privacy) vary in severity across scenarios. Accordingly, this study proposes a risk-tiered principle for the human-AI division of labor: free use in low-risk scenarios, guided use in medium-risk scenarios, and restricted use in high-risk scenarios, with teachers remaining learning designers and final arbiters of assessment. The study offers an integrated framework and practical guidance for rational GenAI use in classroom teaching.

1 Introduction

1.1 Research Background

The release of ChatGPT in November 2022 marked the entry of generative artificial intelligence (GenAI) into mainstream educational practice. Unlike earlier analytical educational AI, which mainly performed recommendation and prediction, GenAI possesses heuristic content generation, conversational context understanding, sequential task execution, and programming language parsing capabilities, enabling it to participate directly in the content production and interaction processes of teaching [1]. Over the past three years and more, GenAI has evolved from a peripheral tool into a routine element of classroom teaching: teachers use it to generate lesson plans and exercises, students use it to obtain instant answers and writing feedback, and schools have begun deploying various subject-specific agents. The speed of this penetration has far outpaced the response of educational institutions and educational research, generating considerable controversy: to what extent has it actually improved teaching, and where does it pose threats?

1.2 Literature Review

Existing research has developed along three main strands.

The first strand focuses on application scenarios. Based on the four core capabilities of

(cc) BY

ChatGPT, Lu et al. [1] were among the first to propose an application framework spanning teacher instruction, learning processes, educational assessment, and academic tutoring, with preliminary validation in exercise generation, automatic problem solving, and assisted grading. The review of 55 studies by Bidry et al. [3] depicted the overall landscape of GenAI applications in education. In K-12 education, two systematic reviews by Zhang et al. [5] and by Lin and Tan [6] mapped application scenarios as well as research goals, activities, and roles; the latter coded 84 empirical studies using Biggs's 3P model and represents an important advance in scenario classification. At the subject level, Su et al. [9] systematically reviewed the pedagogical value and limitations of ChatGPT in mathematics education, while Li et al. [4] traced the evolution of 144 empirical studies in language learning and teaching. Among Chinese-language studies, Lu and Tang [10] constructed a four-level framework of classroom empowerment, namely labor substitution, capability enhancement, human-AI collaboration, and cognitive integration; Gu et al. [11] analyzed the mechanisms, demands, and pathways of GenAI-empowered teaching; and Liu et al. [13] discussed the conceptual foundations and construction challenges of educational large language models.

The second strand concerns actual effectiveness. Evidence in this direction is accumulating rapidly but remains inconsistent. Synthesizing 67 empirical studies, Li et al. [8] found that ChatGPT can support university students' critical and creative thinking when embedded in inquiry-oriented and scaffolded instruction, whereas casual use without instructional design shows no stable gains. The bibliometric and systematic review by Nasr et al. [14] further noted that while GenAI reshapes critical thinking, it blurs the boundary between authentic intellectual effort and AI-assisted performance, complicating assessment design and the interpretation of learning outcomes. Wei [12] summarized the potential and institutional risks based on the experience of American universities. Overall, effectiveness research exhibits scenario-dependent and design-dependent characteristics, yet most reviews merely list findings, and few have systematically stratified evidence strength by scenario type.

The third strand addresses risks and governance. Academic integrity was the earliest concern: the survey of 337 university students by Gruenhagen et al. [15] revealed the prevalence of and attitudes toward using chatbots for assessment assistance, while Lee et al. [16] drew on high-school survey data collected before and after the release of ChatGPT to examine changes in self-reported cheating. Cognitive risks have also received attention: Wang and Qin [18] proposed the concept of thinking dependence, arguing that overreliance on GenAI leads to the degradation of critical thinking, the inhibition of creative thinking, and the weakening of metacognitive ability. Regarding the digital divide, Beckman et al. [17] identified an emerging GenAI divide within the university student population. At the governance level, Liu et al. [19] constructed a multi-stakeholder collaborative governance framework, and Tan [20] analyzed Japan's governance logic of inclusive prudence. However, risk discussions remain largely detached from scenario analysis: existing literature rarely answers the question practitioners care about most, namely which scenarios carry high risk and which carry low risk.

1.3 Research Gap and Research Questions

In summary, the three research strands on scenarios, effectiveness, and risks have each developed rapidly yet remain disconnected: scenario studies do not evaluate evidence, effectiveness studies do not differentiate scenarios, and risk studies are not linked to concrete applications. The absence of an integrated analytical framework leaves frontline teachers and school administrators without directly usable criteria when facing the questions of whether to use GenAI, where to use it, and how to use it. This constitutes the research gap of the present study, which addresses three research questions:

RQ1: What typical application scenarios and tool forms has GenAI formed across classroom teaching processes?

(cc) BY

RQ2: What effectiveness evidence has empirical research reported? For which scenarios is the evidence strong, inconsistent, or merely anticipated?

RQ3: What are the main risks of classroom GenAI application? How can risk levels and human-AI divisions of labor be delineated by scenario?

1.4 Methods and Materials

This study is a systematic literature analysis, with the retrieval and screening process organized with reference to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) reporting guideline (Figure 1).

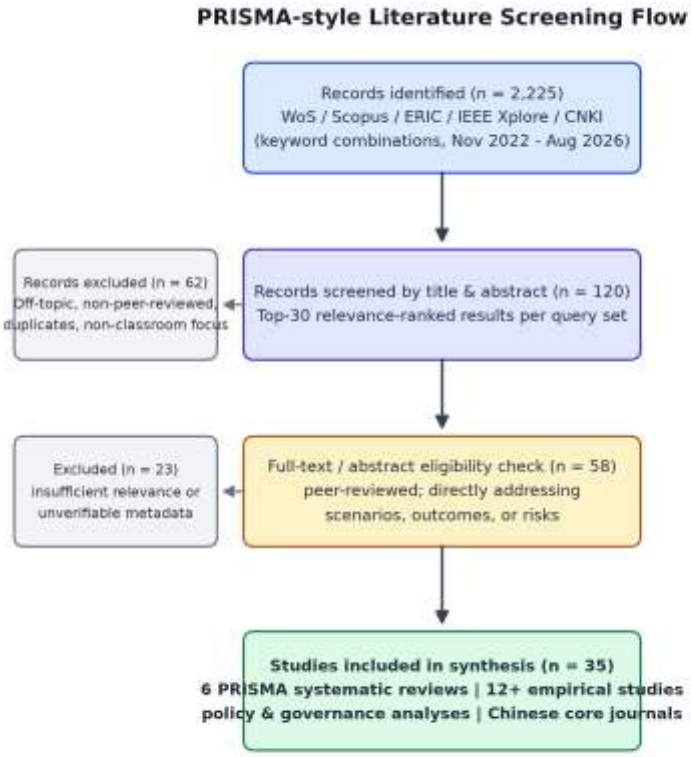


Figure 1. PRISMA-style literature screening flow

(1) Retrieval strategy. In August 2026, four query sets were executed through an aggregated retrieval platform covering Web of Science, Scopus, ERIC, IEEE Xplore, and CNKI: (i) generative AI x education x review; (ii) generative AI x learning outcomes (learning outcomes / academic performance / feedback / tutoring); (iii) generative AI x risks (academic integrity / hallucination / overreliance / digital divide / privacy); and (iv) Chinese-language queries combining generative AI / large language model / ChatGPT with classroom teaching / educational application. The time window spanned November 2022 to August 2026, yielding 2,225 records.

(2) Screening process. The top 30 relevance-ranked records from each query set (120 in total) were screened by title and abstract, and 62 records were excluded for being off-topic, not peer-reviewed, duplicated, or not classroom-focused. The remaining 58 records underwent eligibility assessment (peer-review status, direct relevance to the research questions, and verifiability of metadata), and 23 further records were excluded, leaving 35 included studies. The corpus includes six PRISMA-compliant systematic reviews, more than twelve empirical studies (quasi-experiments, surveys, configurational analysis, and others), and several policy and governance analyses. English-language sources come mainly from SSCI-indexed journals such as Computers and

(cc) BY

Education: Artificial Intelligence, The Internet and Higher Education, and IEEE Transactions on Learning Technologies, while Chinese-language sources come mainly from CSSCI-indexed journals, indicating a generally high quality of the included literature.

(3) Coding and analysis. Framework analysis was adopted: the self-constructed three-dimensional scenario-effectiveness-risk framework served as the coding structure, and the research scenarios, evidence types, key findings, and risk themes of the included studies were categorized and coded, followed by cross-study synthesis to produce the scenario map, the effectiveness stratification, and the risk classification. It should be noted that some empirical data in this study are cited from article abstracts, so the conclusions are directional in nature; this defines the evidential boundary of the present study.

2 Main Part

2.1 Conceptual Definition and Analytical Framework

In this study, generative AI refers to AI systems represented by large language models that can generate new content such as text, code, and images in response to prompts, in contrast to analytical educational AI (such as early adaptive learning platforms) that focuses on prediction and recommendation [1][13]. The key to this definition lies in generation and dialogue: for the first time, technology can participate in instructional interaction as a quasi-agent, which is the root of its profound impact on classroom teaching.

This study constructs a three-dimensional scenario-effectiveness-risk analytical framework (Figure 2). The first dimension maps application scenarios (answering where to use); the second stratifies effectiveness by evidence strength (answering whether it works); and the third identifies and classifies risks (answering where the boundaries lie). The intersection of the three dimensions provides the basis for delineating the human-AI division of labor.

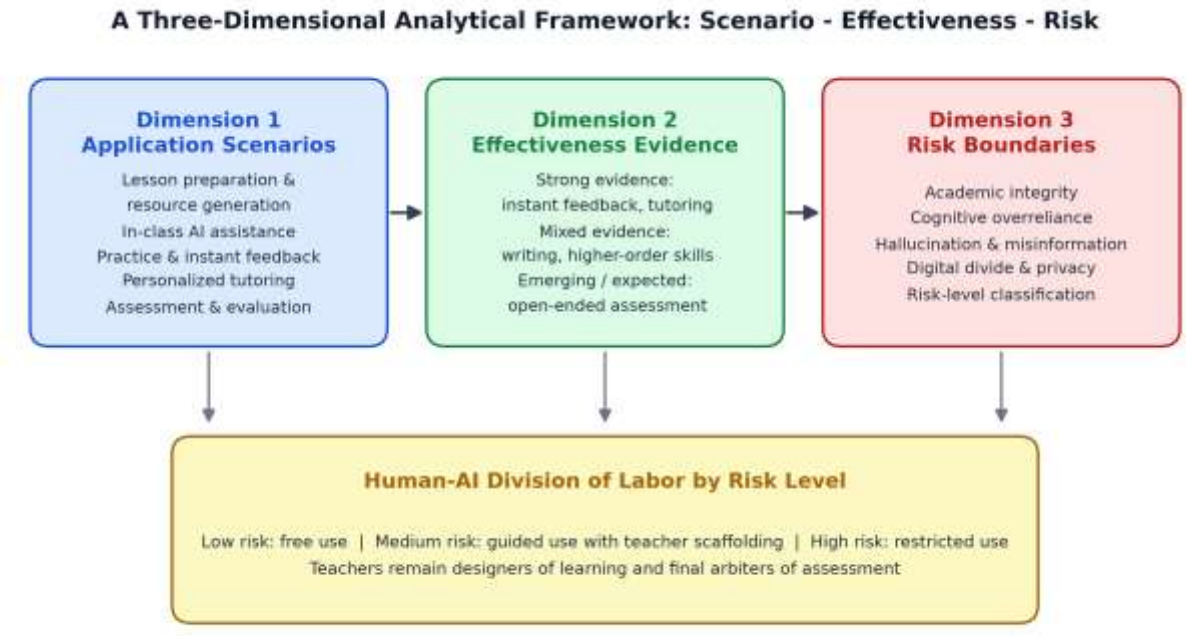


Figure 2. A three-dimensional analytical framework: scenario-effectiveness-risk

2.2 A Map of Application Scenarios (RQ1)

Synthesizing the included literature, GenAI has formed five typical application scenarios in classroom teaching (Table 1).

(cc) BY

Table 1.

Classification of GenAI application scenarios in classroom teaching

Scenario category	Typical tool forms	Representative references
Lesson preparation & resource generation	Lesson plan and syllabus generation; exercise and material generation	[1][23]
In-class teaching assistance	AI teaching assistants; real-time Q&A; dialogic explanation	[10][11]
Practice & instant feedback	Essay grading; programming feedback; hierarchical error correction	[21][22][25]
Personalized tutoring	24/7 subject tutors; adaptive learning paths	[24][27]
Academic assessment	Automated scoring; process-data analytics	[9][14]

(1) Lesson preparation and resource generation. This is the scenario with the highest penetration among teachers. Lu et al. [1] validated the feasibility of ChatGPT in exercise generation and assisted grading. Rutecka et al. [23] prompted four large models to generate curricula and syllabi for bachelor's programs in economics and management, finding that although none of the models matched the completeness of human experts, they could substantially assist course design; notably, their experiment showed that the more specific the prompt, the less accurate the output, revealing the capability boundary of this scenario.

(2) In-class teaching assistance. This scenario takes the form of AI teaching assistants, real-time in-class question answering, and dialogic explanation. The four-level framework of Lu and Tang [10] indicates that classroom empowerment is evolving from labor substitution toward human-AI collaboration; Gu et al. [11] likewise identified coupling pathways between instructional demands and technological affordances at the mechanism level.

(3) Practice and instant feedback. This is currently the most intensively studied scenario. Reddig et al. [21] used GPT-4 within an intelligent tutoring system to diagnose student errors and generate personalized corrective feedback. The SPHERE system developed by Tang et al. [22] supports instructors in generating and human-reviewing personalized feedback at scale for large programming classrooms, forming a mature paradigm of AI generation plus teacher review. Yu et al. [25] conducted a quasi-experiment in interpreter training comparing learners' engagement strategies with AI chatbot feedback versus peer feedback.

(4) Personalized tutoring. Nikolic et al. [24] simulated and evaluated the capability of tools such as ChatGPT-4/4o to serve as 24/7 tutoring systems for engineering and mathematics, while also documenting hallucination risks. The ITS-CAL system developed by Lai and Lin [27] provides immediate, hierarchical feedback for large programming classes.

(5) Academic assessment. Automated scoring and process-data analysis are reshaping assessment, but this scenario is the most controversial. Both the mathematics education review by Su et al. [9] and Nasr et al. [14] noted that distinguishing AI-assisted performance from genuine academic achievement has become a central challenge in assessment design.

2.3 Analysis of Effectiveness Evidence (RQ2)

Stratifying the empirical evidence by strength is the key feature distinguishing this review from general overviews (Table 2).

(cc) BY

Table 2.

Stratification of effectiveness evidence by scenario

Evidence level	Scenarios	Representative empirical findings
Relatively strong	Instant feedback; programming & language learning; personalized tutoring	ITS feedback improves performance [21][27]; positive ChatGPT perception enhances online learning performance via engagement [26]; quasi-experiments report significant learning gains [28][29]
Inconsistent	Writing instruction; higher-order thinking	Gains in critical and creative thinking depend on inquiry-oriented, scaffolded design; casual use shows no stable benefits [8][14]; models underperform humans on complex tasks such as curriculum design [23]
Exploratory	Open-ended assessment; competency-oriented evaluation	AI-assisted performance is hard to distinguish from genuine ability; assessment validity is questionable [9][14]

First, scenarios with relatively strong evidence concentrate on highly structured feedback and tutoring. Lai and Lin [27] found that students with different knowledge levels exhibited divergent behavioral patterns and gains in the intelligent tutoring system, suggesting that effectiveness is not uniformly distributed. The quasi-experiment by Alneyadi and Wardat [28] in a Grade 12 quantum theory course in the United Arab Emirates (experimental group, n = 55) and the quasi-experiment by Jiang et al. [29] in a medical classroom both reported significant improvements in learning outcomes; the latter also found enhanced student interaction, research interest, and critical thinking.

Second, scenarios with inconsistent evidence concentrate on higher-order cognitive goals. The synthesis of 67 empirical studies by Li et al. [8] showed that ChatGPT's support for critical thinking is highly dependent on instructional design. The fsQCA configurational analysis of 29 typical university cases by Liu et al. [30] offers a more fundamental explanation: no single condition (technology, actors, or environment) constitutes a necessary condition for high instructional effectiveness; high effectiveness appears only in two configurations, the actor-technology synergy type and the balanced actor-technology-environment type. This implies that the source of effectiveness is not the technology itself but human-AI collaborative instructional design.

Third, evidence for open-ended assessment scenarios remains exploratory. When GenAI is deeply involved in learning outputs, the question of whether assessment evaluates the student or the model becomes a validity challenge [14]. Reliable empirical support is currently lacking for this scenario, which is precisely where risks are highest (see Section 2.4).

2.4 Risk Identification and Boundary Delineation (RQ3)

The risks reported in the literature can be grouped into five categories, whose severity varies across scenarios.

(1) Academic integrity and assessment failure. Gruenhagen et al. [15] found that students commonly use chatbots to assist with assessment tasks, while the before-and-after high-school surveys of Lee et al. [16] provide rare longitudinal evidence on changes in cheating behavior. Alyafei et al. [31] reviewed institutional and technical countermeasures for safeguarding integrity.

(2) Cognitive overreliance and capability degradation. The integrative review by Noorbehbahani and Oyibo [32] aggregated evidence on the hazards of AI overdependence and cognitive decline together with mitigation strategies. The concept of thinking dependence proposed by Wang and Qin [18] characterizes its educational mechanism: learners actively

(cc) BY

relinquish the exercise of critical, creative, and autonomous thinking.

(3) Hallucination and transmission of erroneous knowledge. Adel and Alani [33] showed that ChatGPT-assisted literature synthesis can reduce workload by approximately 60-65%, but with highly variable accuracy; Nikolic et al. [24] likewise documented hallucination problems in tutoring scenarios. Knowledge accuracy risks are particularly salient in science question-answering and factual explanation scenarios.

(4) The digital divide and educational equity. Beckman et al. [17] revealed a GenAI divide within the student population, implying that the technology may amplify rather than bridge existing gaps.

(5) Data privacy and the protection of minors. Classroom scenarios involve large volumes of learning data from minors, which Liu et al. [19] listed among six core ethical concerns.

Based on the above analysis, this study proposes a scenario-based risk classification and human-AI division-of-labor boundary table (Table 3) as the answer to RQ3 and the core practical contribution of this study.

Table 3.

Risk classification and human-AI division of labor by scenario

Risk level	Scenarios	Usage principle	Human-AI division of labor
Low (free use)	Practice & instant feedback; personalized tutoring; preparation material generation	Encouraged, with routine monitoring	AI generates and pre-judges; teachers spot-check and backstop
Medium (guided use)	In-class assistance; writing instruction; higher-order thinking tasks	Used under teacher-designed scaffolding	Teachers design tasks and prompt boundaries; AI provides scaffolds; teachers assess the process
High (restricted use)	Open-ended assessment; summative evaluation; unguarded free generation	Restricted, or assessment format changed	Teachers lead assessment design; AI assists analysis only and must not be the final basis for grading

2.5 Discussion: Implications for Teaching Practice

The core conclusion emerging from the intersection of the three dimensions is that there is no blanket answer to whether GenAI should be used in classrooms; there are only scenario-specific answers about where and how to use it. Low-risk scenarios, where effectiveness evidence is strongest, should be actively promoted. High-risk scenarios are precisely where evidence is weakest and risks most concentrated; they should be addressed through reform of assessment formats (such as process-based assessment, oral defense, and in-class completion) rather than blanket prohibition [14][31].

Institutional support is equally critical. Policy analyses of American flagship universities show that instructional decisions on GenAI use are commonly delegated to the course level [34][35], while Japan's experience of inclusive prudence demonstrates the feasibility of reserving pilot space for innovation within ethical regulation [20]. For teachers, the immediate priority is to develop competencies for human-AI collaborative instructional design: the configurational evidence of Liu et al. [30] indicates that the synergy between teacher agency and technology is the key driver of effectiveness, and Wang and Qin [18] further propose that teachers transform into creators of cognitive conflict to counter thinking dependence. At the school level, scenario-specific usage guidelines,

(cc) BY

teacher training systems, and new academic integrity rules are needed.

3 Conclusion

Addressing the fragmentation among scenario, effectiveness, and risk research on classroom GenAI application, this study conducted a systematic literature analysis following the PRISMA protocol and constructed an integrated three-dimensional framework, reaching three conclusions. First (RQ1), GenAI has formed stable application patterns in five scenarios: lesson preparation and resource generation, in-class teaching assistance, practice and instant feedback, personalized tutoring, and academic assessment, with the richest empirical accumulation in practice feedback and personalized tutoring. Second (RQ2), effectiveness evidence is significantly stratified: relatively strong for structured feedback and tutoring scenarios, inconsistent and highly design-dependent for higher-order thinking scenarios, and insufficient for open-ended assessment; configurational analysis further indicates that effectiveness stems from human-AI collaboration rather than the technology itself. Third (RQ3), the five categories of risk, namely academic integrity, cognitive overreliance, hallucination, the digital divide, and data privacy, vary in severity across scenarios, on the basis of which a three-tier human-AI division of labor, namely free use, guided use, and restricted use, can be delineated.

This study has limitations: it relies on literature analysis, its conclusions are directional, some empirical data are cited from abstracts, and no first-hand classroom empirical work was conducted; the scenario classification is mainly inductive from existing studies and may evolve with technological development. Future research can be deepened in three directions: conducting scenario-specific classroom quasi-experiments to directly test the predictive power of the three-dimensional framework; tracking the adaptability of risk-tiered boundaries across educational stages and disciplines; and examining pathways for cultivating teachers' human-AI collaborative design competencies.

References

1. Lu Yu, Yu Jinglei, Chen Penghe, et al. Educational applications and prospects of generative artificial intelligence: A case study of the ChatGPT system. China Distance Education, 2023, No.4, pp.24-31,51. DOI: 10.13541/j.cnki.chinade.20230301.001
2. Li Sen, Zheng Lan. Challenges posed by generative artificial intelligence to classroom teaching and countermeasures. Curriculum, Teaching Material and Method, 2024, Vol.44, No.1, pp.39-46. DOI: 10.19877/j.cnki.kcjcjf.2024.01.003
3. Bidry M., Hanine M., Ouaguid A., et al. Transforming Education With Generative AI: A Comprehensive Review of Advancements, Challenges, and Future Opportunities. IEEE Access, 2025, Vol.13, pp.202938-202955. DOI: 10.1109/ACCESS.2025.3636891
4. Li B., Tan Y.L., Wang C., et al. Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. Computers and Education: Artificial Intelligence, 2025, Vol.9, Art.100445. DOI: 10.1016/j.caeai.2025.100445
5. Zhang T., Lai Y.C., Yu P.L.H. Generative artificial intelligence in K-12 education: A systematic review. Research and Practice in Technology Enhanced Learning, 2026, Vol.21. DOI: 10.58459/rptel.2026.21034
6. Lin X., Tan H. A Systematic Review of Generative AI in K-12: Mapping Goals, Activities, Roles, and Outcomes via the 3P Model. Systems, 2025, Vol.13, No.10, Art.840. DOI: 10.3390/systems13100840
7. Ogunleye B., Zakariyyah K.I., Ajao O. A Systematic Review of Generative AI for Teaching and Learning Practice. Education Sciences, 2024, Vol.14, No.6, Art.636. DOI: 10.3390/educsci14060636
8. Li C., Cui H., Hagedorn L.S. The cognitive impact of ChatGPT in higher education: A systematic review of critical and creative thinking outcomes. Computers and Education: Artificial Intelligence, 2026, Vol.10, Art.100571. DOI: 10.1016/j.caeai.2026.100571
9. Su C.S., Hsu C.C., Campos J.A.G., et al. ChatGPT As a Resource to Strengthen Mathematics Teaching and Learning: A

(cc) BY *dtai.uz*

Systematic Review. Journal of Computer Assisted Learning, 2026, Vol.42, No.3. DOI: 10.1002/jcal.70241
10. Lu Yu, Tang Xiaoyu. Hierarchical forms and advancement paths of generative AI empowering classroom teaching. E-education Research, 2025, Vol.46, No.6, pp.75-82,106. DOI: 10.13811/j.cnki.eer.2025.06.010
11. Gu Xiaoqing, et al. Mechanisms, demands, and pathways of generative AI empowering teaching. Journal of the Chinese Society of Education, 2025, No.4, pp.15-22.
12. Wei Shunping. Potential and risks of ChatGPT in higher education: Experience and implications from American universities. Modern Distance Education, 2024, No.3, pp.18-27. DOI: 10.13927/j.cnki.yuan.20240628.003
13. Liu Ming, Wu Zhongming, Yang Xiao, et al. The connotation, construction, and challenges of educational large language models. Modern Distance Education Research, 2024, Vol.36, No.5, pp.50-60. DOI: 10.3969/j.issn.1009-5195.2024.05.006
14. Nasr N.R., Tu C.-H., Sujo-Montes L., et al. Generative artificial intelligence impact on critical thinking in higher education: a comprehensive bibliometric analysis and systematic review. Journal of Computers in Education, 2026. DOI: 10.1007/s40692-026-00396-z
15. Gruenhagen J.H., Sinclair P.M., Carroll J.-A., et al. The rapid rise of generative AI and its implications for academic integrity: Students' perceptions and use of chatbots for assistance with assessments. Computers and Education: Artificial Intelligence, 2024, Vol.7, Art.100273. DOI: 10.1016/j.caeai.2024.100273
16. Lee V.R., Pope D., Miles S., Zarate R.C. Cheating in the age of generative AI: A high school survey study of cheating behaviors before and after the release of ChatGPT. Computers and Education: Artificial Intelligence, 2024, Vol.7, Art.100253. DOI: 10.1016/j.caeai.2024.100253
17. Beckman K., Apps T., Howard S.K., et al. The GenAI divide among university students: A call for action. The Internet and Higher Education, 2025, Vol.67, Art.101036. DOI: 10.1016/j.iheduc.2025.101036
18. Wang Tianping, Qin Peng. The risk of thinking dependence in educational applications of generative AI and its avoidance. Modern Distance Education Research, 2026, Vol.38, No.3, pp.39-46. DOI: 10.3969/j.issn.1009-5195.2026.03.005
19. Liu Bangqi, Huang Qianqian, Zhang Jinxia. A multi-stakeholder collaborative governance framework and practical strategies for ethical risks of GenAI educational applications. Modern Distance Education Research, 2025, Vol.37, No.5, pp.48-58. DOI: 10.3969/j.issn.1009-5195.2025.05.006
20. Tan Jianchuan. Inclusive prudence: An analysis of Japan's governance logic for educational applications of generative AI. Comparative Education Review, 2026, Vol.48, No.3, pp.3-11,52. DOI: 10.20013/j.cnki.ICE.2026.03.01
21. Reddig J.M., Arora A., MacLellan C.J. Generating In-Context, Personalized Feedback for Intelligent Tutors with Large Language Models. International Journal of Artificial Intelligence in Education, 2025, Vol.35, No.6, pp.3459-3500. DOI: 10.1007/s40593-025-00505-6
22. Tang X., Wong S., Huynh M., et al. SPHERE: Supporting Personalized Feedback at Scale in Programming Classrooms with Structured Review of Generative AI Outputs. Proceedings of the ACM, 2025. DOI: 10.1145/3706599.3720203
23. Rutecka P., Cicha K., Rizun M., et al. Generative AI in Curriculum Design: Empirical Insights Into Model Performance and Educational Constraints. IEEE Transactions on Learning Technologies, 2025, Vol.18, pp.757-768. DOI: 10.1109/TLT.2025.3587081
24. Nikolic S., Vu B.A., Di Y., et al. Generative Artificial Intelligence (GenAI) 24x7 Tutor: A simulation of the capability of ChatGPT, Wolfram GPT and Tutor Me GPT to accurately and effectively tutor engineering and math content. STEM Education, 2026,

(cc) BY *dtai.uz*

Vol.6, No.1, pp.56-83. DOI: 10.3934/steme.2026004
25. Yu Y., Wei W., Chen Z. Comparing learners' engagement strategies with feedback from a Generative AI chatbot and peers in an interpreter training programme: a quasi-experimental study. The Interpreter and Translator Trainer, 2025, Vol.19, No.3-4, pp.338-356. DOI: 10.1080/1750399X.2025.2533069
26. Mehrvarz M., Salimi G., Abdoli S., et al. How does students' perception of ChatGPT shape online learning engagement and performance? Computers and Education: Artificial Intelligence, 2025, Vol.9, Art.100459. DOI: 10.1016/j.caeai.2025.100459
27. Lai C.-H., Lin C.-Y. Analysis of Learning Behaviors and Outcomes for Students with Different Knowledge Levels: A Case Study of Intelligent Tutoring System for Coding and Learning (ITS-CAL). Applied Sciences, 2025, Vol.15, No.4, Art.1922. DOI: 10.3390/app15041922
28. Alneyadi S., Wardat Y. Integrating ChatGPT in Grade 12 Quantum Theory Education: An Exploratory Study at Emirate School (UAE). International Journal of Information and Education Technology, 2024, Vol.14, No.3, pp.398-410. DOI: 10.18178/ijiet.2024.14.3.2061
29. Jiang Fengguang, Zhang Qinglin, Wu Xinyi, et al. Systematic reconstruction of medical classroom teaching driven by generative AI: A quasi-experimental study from the perspective of constructive alignment. Modern Educational Technology, 2026, Vol.36, No.1, pp.84-93. DOI: 10.3969/j.issn.1009-8097.2026.01.009
30. Liu Bangqi, Huang Qianqian, Liu Mengdie, et al. Configurational paths and practical strategies of factors influencing the effectiveness of GenAI teaching applications: An analysis of 29 typical university cases. Modern Educational Technology, 2026, Vol.36, No.4, pp.53-62. DOI: 10.3969/j.issn.1009-8097.2026.04.006
31. Alyafei R.A.S.M., Al-Kfairy M., Alfandi O., et al. Safeguarding students' academic integrity in the age of generative AI: challenges and solutions. Kybernetes, 2026. DOI: 10.1108/K-04-2025-0878
32. Noorbehbahani F., Oyibo K. AI-overdependence and human cognitive decline: Hazards, evidence, and mitigation strategies. Computers in Human Behavior Reports, 2026, Vol.22, Art.101102. DOI: 10.1016/j.chbr.2026.101102
33. Adel A., Alani N. Can generative AI reliably synthesise literature? exploring hallucination issues in ChatGPT. AI & Society, 2025, Vol.40, No.8, pp.6799-6812. DOI: 10.1007/s00146-025-02406-7
34. LaFrance J. Governing Generative Artificial Intelligence: Institutional Policies and Guidelines at America's Flagship Universities. Educational Policy, 2026. DOI: 10.1177/08959048261450995
35. An Y., Yu J.H., James S. Investigating the higher education institutions' guidelines and policies regarding the use of generative AI in teaching, learning, research, and administration. International Journal of Educational Technology in Higher Education, 2025, Vol.22, No.1, pp.1-23. DOI: 10.1186/s41239-025-00507-3

(cc) BY