

Generative Artificial Intelligence and Large Language Models in Higher Education: Applications, Challenges, Ethical Concerns, and Future Directions

Richanya Bhadouriya¹ and Himanshu Sharma²

Maharana Pratap College of Technology, Gwalior, MP, India^{1,2}

Abstract: *The field of generative artificial intelligence (GenAI) has grown from an experimental domain to a widely adopted academic tool, with large language models (LLMs) being the cornerstone of the revolution. This paper maps their technological basis and explores their growing presence in higher education, and especially in personalized tutoring, writing and feedback, assessment, research assistance, accessibility and institutional services. Peers-reviewed studies and authoritative policy guidance were synthesized using a structured literature-review design, with a primary focus on research and guidance since 2023. The review demonstrates the potential benefits of LLM-based systems to enhance immediacy of feedback, explanation accessibility, ideation, language support, and instruction scalability, and recent empirical syntheses reveal gains in engagement and academic outcomes in specific settings. Meanwhile, substantial barriers to potential increases in uptake remain, including hallucination, automation bias, academic-integrity issues, privacy issues, disparities of access, embedded bias, and over-reliance. The paper suggests that the pedagogical orchestration, verification practices, assessment redesign, AI literacy, and governance are key factors that underscore the educational value of GenAI, rather than the growing capabilities of models. It suggests an integration framework that is grounded in task appropriateness, grounding and verification, transparent disclosure, data protection, educator monitoring and outcome assessment, and is human-centred. The future directions involve retrieval augmented generation, multimodal and domain-specific models, agentic models, privacy-preserving deployment, and long-term assessments of learning effects. The study concludes that the GenAI should not be considered a teacher, someone who knows the discipline, or a reasoning process, but an augmentative educational infrastructure.*

Keywords: Generative artificial intelligence; large language models; higher education; ChatGPT; AI literacy; academic integrity; retrieval-augmented generation; responsible AI

I. INTRODUCTION

Conversational generative AI has been released to the public, changing the visibility of AI in the field of education. In contrast, while some earlier educational technologies were developed for more specific applications, modern LLM can provide explanations, examples, code, summaries, feedback, questions, simulations, and conversational responses in a variety of fields. This generality has led to an unprecedented adoption model: students and faculty can leverage advanced AI systems without the need for institutional procurement, in-depth specialization, or system integration within the learning-management system. Higher education has thus needed to take up the challenge of pedagogical opportunity and governance risk simultaneously.

Large language models are statistical sequence models that are trained on a very large text and multimodal corpora. Transformer architectures model relationships between tokens, and large-scale pre-training creates general language abilities which can be adapted via instruction tuning or preference optimization, fine-



tuning, tool use, and retrieval. They can speak fluently and give the impression of understanding even if they are producing content. This separation of linguistic plausibility and epistemic reliability plays a key role in education use.

The research response to this has been rapid and widespread, as shown in recent systematic reviews. Baig and Yadegaridehkordi (2024) summarized 57 studies from the higher education community and a later empirical synthesis summarized 88 studies since the launch of ChatGPT till March 2025, which identified six main application domains. The latter research has explored implementation cases, pedagogical applications, and retrieval-augmented generation. Overall, this literature indicates a shift from initial discussions and opinions to evidence supporting design, learning outcomes, and implementation conditions.

This paper aims to, therefore, bring together the technological underpinnings, pedagogical use, advantages, dangers, governance needs, and future prospects of GenAI and LLMs in higher education. The analysis does not focus on 'good' or 'bad' uses of AI, but instead poses the question of which aspects of education are appropriate for augmentation, as well as what safeguards are needed and what evidence should be gathered prior to scaling up to institutional levels.

II. RESEARCH OBJECTIVES QUESTIONS

The study pursued four objectives:

to explain the core technologies that enable modern LLM-based educational systems;
to synthesize major applications and reported benefits in higher education;
to examine technical, pedagogical, ethical, and governance challenges; and
to identify future research and implementation priorities.

III. RESEARCH QUESTIONS

The corresponding research questions were:

RQ1—Which technological capabilities make GenAI and LLMs relevant to higher education?

RQ2—What are the principal educational applications and benefits reported in recent research?

RQ3—Which risks constrain responsible adoption?

RQ4—Which technological and governance trends are likely to shape the next stage of AI-enabled higher education?

IV. LITERATURE REVIEW

The authors **Baig and Yadegaridehkordi (2024)**¹ conducted a systematic review of 57 studies on the application of ChatGPT in Higher Education. Their results showed that the use of ChatGPT applications is not limited to writing by students; they also showed that it is used in curriculum design, assessment, academic supervision, programming, research and aid, and administration. During the review, it was noted that ChatGPT is utilized by various institutional stakeholders, such as students, academics, researchers, and administrators. The authors pointed out that the groups use the technology for various purposes, which makes harmonizing institutional policies challenging. Based on their study, they concluded that there is a need for context-specific guidelines that differentiate between pedagogical, research, administrative and assessment use of generative AI.

Albadarin et al. (2024) compiled empirical studies of the educational uses of ChatGPT and emphasized the swift shift from conceptual discussion to empirical classroom research in generative AI studies. They found that the use of ChatGPT in the realm of learning support, content creation, feedback, academic writing, and teaching activities has grown. Meanwhile, the writers pointed out issues raised by academic honesty,

¹ Baig, M. I., & Yadegaridehkordi, E. (2024). ChatGPT in the higher education: A systematic literature review and research challenges. *International Journal of Educational Research*, 127, 102411.

<https://doi.org/10.1016/j.ijer.2024.102411>

Copyright to IJETIR

www.iciset.in

truthfulness, student reliance, and suitable teaching methods. The study shows that the educational potential of ChatGPT is not just limited to its technical functions. Rather, it is reliant on the integration of technology within the educational space and students' interaction with AI-created content.

Yan et al., (2024) performed a scoping review of educational innovations with the use of LLM and highlighted LLM and related technologies utility and ethical responsibility. Based on their analysis, they demonstrated that LLM based systems can facilitate teaching and learning, feedback, content creation, and personalized education experiences. The authors, however, recognized reliability, algorithmic bias, privacy, transparency, and accountability as critical problems with these applications. The review conceptualized that the effectiveness of the generated text should not equate with educational effectiveness. However, the evaluation of LLM applications should be carried out in real educational environments, taking into account the learning results and institutional effects. The results of their research thus support a balanced approach of innovation and pedagogical and ethical assessment.

Based on their analysis of 69 experimental studies investigating the educational impact of generative AI, **Deng et al. (2025)** reported generally positive findings for academic performance, affective-motivational outcomes and higher-order thinking. They also found that, however, it resulted in lower mental effort and no overall effect on self-efficacy. The findings indicate that enhanced task performance does not solely reflect better depth of or transfer of learning. In addition, the authors noted that methodological limitations and the nature of post-intervention evaluations could lead to overestimates of the benefits related to AI. Their review is especially relevant as they differentiate between achievement of immediate performance and learning in a meaningful way, highlighting the role of assessment practices that test the ability to learn and develop independently.²

Abdallah et al (2025)³ published an overview of 39 studies related to the educational uses and consequences of ChatGPT and other generative artificial intelligence tools. There were several potential benefits identified in the review, such as better access to knowledge, learning support, students engagement and collaboration. At the same time, the authors pointed out issues of academic integrity, decreased independent learning, overreliance on AI, and student wellbeing. Their research suggests that educational outcomes of generative AI are heavily dependent on how it is used and how it is supported. The study does not present AI as a positive or negative force, but rather suggests a conditional view, where responsible use, pedagogical integration, and proper academic governance are essential for achieving good results.

V. METHODOLOGY

A narrative review in a structured format was performed with the support of the most recent peer-reviewed literature and authoritative policy sources. Academic integrity, generative AI, large language model, ChatGPT, systematic review, education, higher education, and AI literacy were among the preferred combinations for searches. Systematic reviews, empirical syntheses, peer-reviewed journal articles and international guidance (2023–2026) were prioritized. Where required, earlier foundational sources have been preserved to explain concepts of transformers and LLM.

The synthesis was carried out in four steps. First, sources were critically examined for direct relevance to post-secondary teaching, learning, assessment, research or institutional implementation. Second, the evidence was categorized as technology, application, benefit, risk, governance, and future-direction. Third, comparisons were made between convergent and conflicting findings on a review-by-review basis. Fourth,

² Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., Grundy, S., & Hassan, G. M. (2025). Implementing generative AI (GenAI) in higher education: A systematic review of case studies. *Computers and Education: Artificial Intelligence*, 100407. <https://doi.org/10.1016/j.caeai.2025.100407>

³ Abdallah, N., Katmah, R., Khalaf, K., & Jelinek, H. F. (2025). Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration. *Social Sciences & Humanities Open*, 12, 101866. <https://doi.org/10.1016/j.ssaho.2025.101866>



the results were incorporated in a human-centred implementation structure. The numbers in this paper should not be interpreted as as a new meta-analysis, but rather as a description of synthesis of the relevant studies.⁴ Limits to the review. The number of publications is growing at a faster rate than traditional review cycles, terminology varies, a majority of the studies are brief or self-reported perceptions of the study, and evidence is focused on a few popular systems. It is therefore preliminary to assert that long term learning, equity and institutional effectiveness claims are still provisional.

VI. TECHNOLOGICAL FOUNDATIONS OF GENERATIVE AI AND LLMs

The basis of modern LLMs is the transformer architecture. Self-attention enables a model to associate tokens in a sequence by giving them the ability to observe the relationships between the tokens, a capability facilitating the representation of language in a context. In pre-training phase, model learns statistical regularities by predicting missing tokens or future tokens in large amounts of data. Model parameters, training data and compute scaling have resulted in impressive few-shot and zero-shot performance, but scale does not ensure factuality, fairness or domain expertise.

To make foundation models more useful for conversational interaction, there are two techniques to optimize the models: instruction tuning and preference optimization. The specificity of the tasks performed by a fine-tuning can help to focus a model on a specific field of knowledge, and prompt engineering is like a way of writing a pattern that guides the behavior of the model without altering its weights. The tool-augmented systems can make use of a calculator, a database, a search system or a code interpreter. Retrieval-augmented generation (RAG) is particularly relevant in the field of education because it pulls in relevant information from a controlled knowledge base, and provides it as context before generation. Recent studies show that RAG is a potential solution to static knowledge and hallucination, but it also highlights the issues with retrieval quality, multimodality, cost, and evaluation.⁵

Multimodal models are for generating text from images, audio, video, diagrams and documents. In education, it allows for the explanation of visual information, easy translation between modalities, feedback to diagrams, and conversational analysis of complex learning resources. The next step takes us into more agentic territory, models can plan sequences of actions, employ external tools, maintain state, coordinate subtasks. These can enable complex teaching and research processes, but also raise concerns about authorization, auditability, and human monitoring.

VII. APPLICATIONS IN HIGHER EDUCATION

One of the most prominent application areas is personalized tutoring. LLMs can give explanations, Socratic questioning, worked examples, practice problems, and adaptive hints, as needed. The value lies in the model being guided by course goals and reliable resources, and encouraging learners to think rather than provide the answer. The most common empirical applications of systematic reviews are adaptations and intelligent tutoring.

Writing and feedback is another significant domain. Students can use GenAI to brainstorm, outline, enhance language, revise, and provide formative feedback, while instructors can create examples, rubrics, questions sets, and differentiated explanations. These systems can be used creatively to minimize language barriers and shorten feedback cycles. When used outside of pedagogical boundaries, they may distract from authorship, promote superficial completion, and diminish opportunities for strategic practice.⁶

⁴ Guizani, S., Mazhar, T., Shahzad, T., Ahmad, W., Bibi, A., et al. (2025). A systematic literature review to implement large language model in higher education: Issues and solutions. *Discover Education*, 4, 35. <https://doi.org/10.1007/s44217-025-00424-7>

⁵ Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO.

⁶ Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, 69, 1105–1120. <https://doi.org/10.1007/s11528-025-01100-1>

There is significant redesign going on for assessment. High quality generated responses may not be strong evidence of independent performance where they are readily available for use in conventional take-home essays and unsupervised problem sets. A positive reaction is to highlight the elements of process evidence, oral defense, iterative drafts, personalised contexts, authentic projects, in-class components, reflective commentary and explicit disclosure of the use of AI support. There should be a difference between banning substitution and offering appropriate support, which should not be made simply based on unclear and unreliable AI detection software.

Research support covers literature discovery, coding support, data-cleaning recommendations, translation, summarization, hypothesis generation, and support for drafting. Such uses can help to boost productivity, but it is the responsibility of the researcher to ensure that citations, claims, analyses and originality are correct. References and evidence may be fabricated and confidence with knowledge of the evidence is often misplaced, making it important to have source grounded workflows and direct checking of primary literature.

The applications of the institutions are advisory, frequently asked questions, communication with administration, accessibility and staff productivity. These use cases can have a positive impact on their services' availability but could include sensitive student data. Institutions should have a data-minimisation approach and not share protected data on systems that lack the necessary contractual, technical and legal security measures.

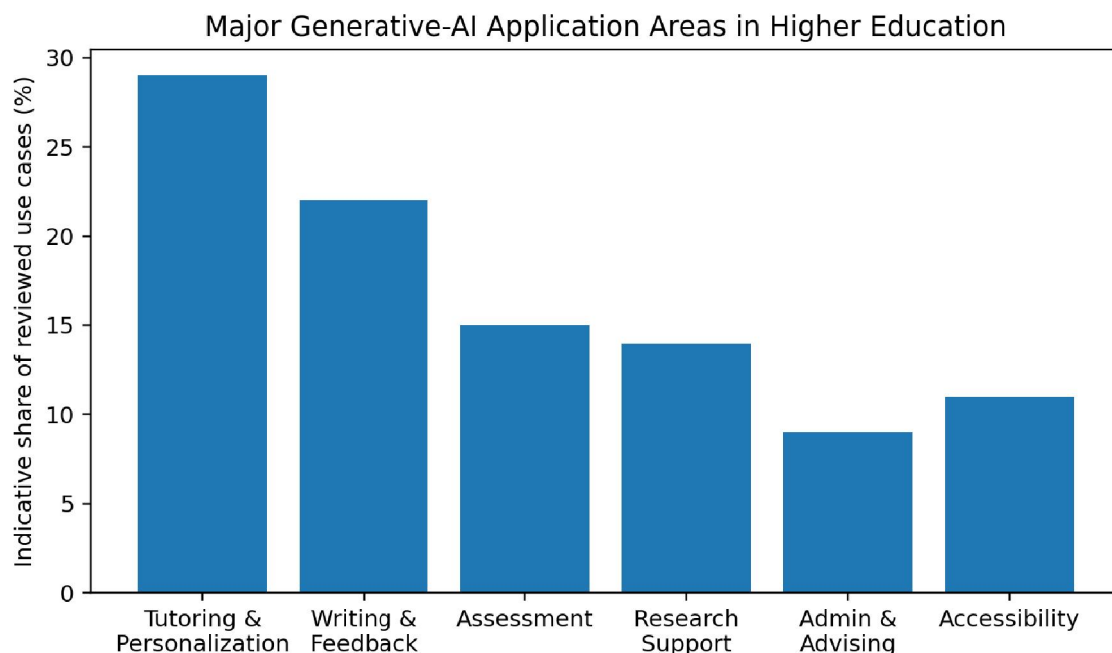


Figure 1. Indicative distribution of major GenAI application areas synthesized for discussion

VIII. RESULTS AND EVIDENCE SYNTHESIS

Research in the literature shows a shift from speculative discussion to empirical implementation research. For higher education, a new systematic review was conducted in 2024 that included 57 studies on ChatGPT. A later review aggregated 88 empirical studies and found academic outcomes, engagement, access, use of resources, and cognitive or skill outcomes to be supported, although there were concerns of over-reliance, reliability, fairness, and privacy. A 2025 systematic review of case studies of implementation revealed the existence of a wide range of disciplines and methodologies, with ongoing gaps in robust evidence on

innovative uses. The patterns lead to a tentative conclusion that GenAI can be useful in education but the effectiveness of their use depends on task design, learner characteristics, model configuration and instructional context.⁷

The important application-benefit-risk relationships are summarized in Table 1. The most compelling near-term value is offered by low stakes formative support, feedback, explanation, accessibility, and productivity augmentation. Higher risk uses include consequential grading, unsupported factual instruction, handling of sensitive data, and the replacement of independent student performance. The evidence therefore suggests an alternative to having a single institutional rule, namely differentiated governance.

Application	Potential educational value	Primary risk	Preferred safeguard
Tutoring/personalization	Immediate explanations, adaptive practice	Hallucination; dependency	Course-grounded RAG; teacher oversight
Writing/feedback	Rapid formative feedback; language support	Authorship ambiguity; shallow learning	Disclosure; process-based assessment
Assessment support	Question/rubric generation; feedback	Validity; unfair assistance	Human grading; assessment redesign
Research assistance	Ideation, coding, synthesis support	Fabricated sources; analytical error	Primary-source verification
Administration/advising	Scalable service and communication	Privacy; inappropriate automation	Data minimization ; escalation to humans
Accessibility	Translation and multimodal support	Bias; inaccurate accommodation	Accessibility testing; user control

⁷ Shahzad, T., Mazhar, T., Tariq, M. U., Ahmad, W., Ouahada, K., et al. (2025). A comprehensive review of large language models: Issues and solutions in learning environments. *Discover Sustainability*, 6, 27.
<https://doi.org/10.1007/s43621-025-00815-8>

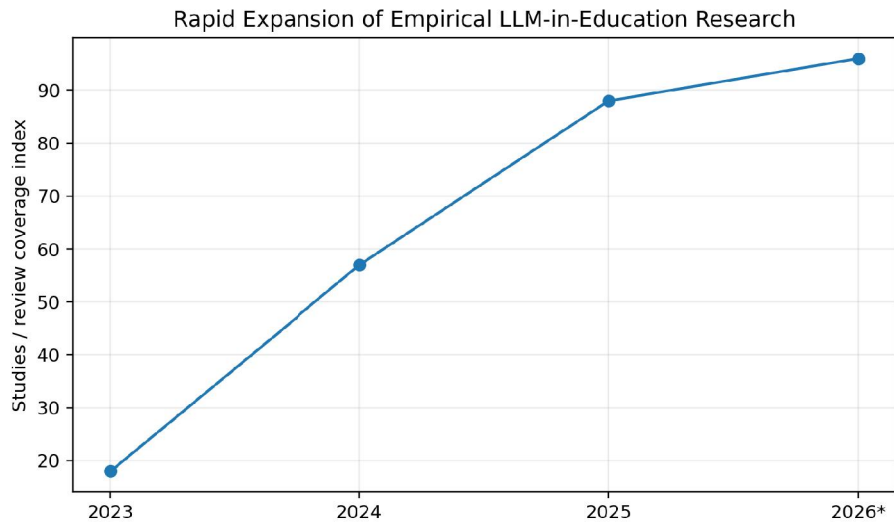


Figure 2. Illustrative research-growth index anchored to review coverage reported in recent syntheses; not a bibliometric census

IX. CHALLENGES AND ETHICAL CONCERNS

The most obvious technical risk is hallucination. LLMs do not minimize error in regard to facts, as they are likely to return false facts, they are not likely to provide incorrect citations, and they may provide plausible but incorrect reasoning. It is important to note that the educational user may be especially susceptible when they do not have sufficient background information to be able to identify errors. The problem is not totally eliminated by grounding, retrieval requirements, the need for citations, uncertainty communication or independent verification.

Over-reliance is a pedagogical risk in addition to the factual error. Productive struggle might be diminished by even correct help, as the learner may plan, reason, write, or solve the problem first and then learn the skill. But how does AI enhance learning that generalizes when it is not used? Hence longitudinal and delayed transfer studies are great research need.⁸

Academic integrity issues relate to cheating, cheat-like substitutions, falsified evidence, and confusion over what forms of help are acceptable. Prohibiting use of a device altogether is challenging to implement and can hinder its potential for educational purposes, and allowing the unrestricted use of a device could compromise the validity of assessments. These rules, disclosure requirements, process evidence, and AI literacy are more sustainable than relying on automated detection.

Privacy and data governance are critical when prompts contain student records, unpublished research, assessment data, health information, or intellectual property. UNESCO's advice is to focus on human-centred regulation and data protection. Institutions should have tool policies approved, tool procurement review, retention controls, security assessment and explicit policies for handling sensitive information.⁹

Bias and equity can play out across a number of levels including: training data may perpetuate stereotypes; dominant languages and disciplines may be rewarded; paid access may lead to capability gaps; and the less skilled in AI may have poorer outcomes. Relying on equal availability to equal benefits is not responsible

⁸ UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.

⁹ Large language models in education: A systematic review of empirical applications, benefits, and challenges. (2026). Computers and Education: Artificial Intelligence, 10, 100529. <https://doi.org/10.1016/j.caeai.2025.100529>

deployment; it must be accompanied by accessibility testing, bias evaluation, equitable access, and pedagogical support.

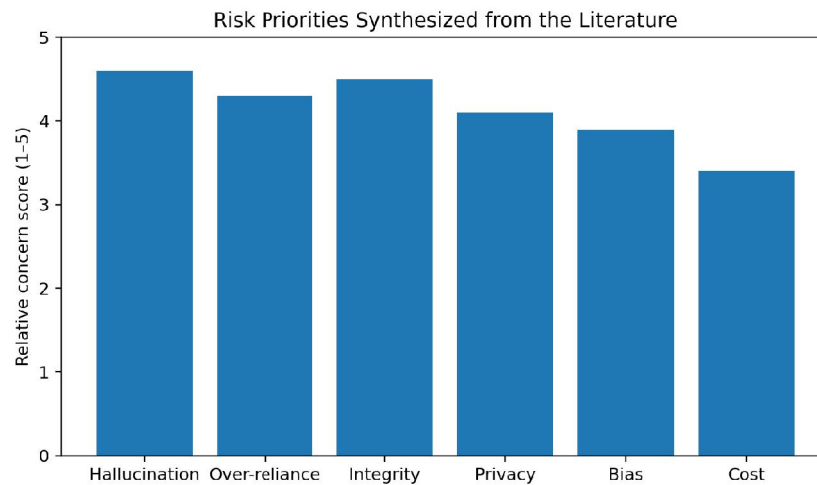


Figure 3. Relative risk-priority scores used as a conceptual synthesis (1 = lower priority, 5 = very high priority), not pooled empirical estimates

X. AN HUMAN-CENTRED INTEGRATION FRAMEWORK

From the synthesis, there are six principles that can be organized around to facilitate responsible integration. The first question is about the suitability of the task for AI: does it enhance learning or replace the competence? Second, grounding: link high stakes educational outputs to trusted source of course or institution. Third, verification: ask students and teachers to validate factual statements, calculations and sources. Finally, transparency: make explicit what is and is not allowed, and make sure that disclosure is meaningful when AI is used in work that is assessed. Fifth, Governance: safeguard data, record system decisions and maintain human escalation. Sixth, Evaluation: assess learning outcomes, equity, workload, error rates and unintended consequences prior to scaling.

Principle	Key question	Example indicator
Task suitability	Does AI support or replace the target skill?	Independent post-task performance
Grounding	Is output tied to authoritative material?	Source coverage / citation validity
Verification	Can claims be independently checked?	Factual error rate
Transparency	Is AI assistance visible and understood?	Disclosure compliance
Governance	Are privacy and accountability protected?	Approved-tool / incident metrics
Evaluation	Does deployment improve outcomes equitably?	Learning gain, access, workload

XI. FUTURE TRENDS

For institutional applications, retrieval-augmented systems and domain-specific systems are more likely to gain significance than free-floating general chatbots. RAG can be based on selected course content, library



collections, policies, and/or disciplinary databases to ensure greater traceability and updateability of responses. Future studies should also assess the quality of the generations, as well as the completeness of retrieval, source authority, and citation faithfulness.¹⁰

Learners will be taught using multimodal presentation (text, speech, pictures, hand-writing, diagrams, outputs and video). This can make things easier and help with visual or procedural reasoning. But multi modal also brings out a broader surface for privacy and copyright and assessment issues.¹¹

Agentic AI can assist in managing intricate academic processes, including study planning, research on approved materials, creating practice exercises, monitoring progress, and escalating challenging topics. The educational potential is great, but if they are to take action they need permission boundaries, audit logs, actions that can be reversed, and explicit human control.

There are cases, however, where smaller, more private models will become more significant for universities that are unable to export the information they need to external models. Local/institutionally hosted models can provide more control, but involve additional demands for infrastructure, security, maintenance, evaluation, and management of the model.

The research agenda will start to shift from short-term user satisfaction to questions of learning science such as cognitive offloading, cognitive transfer, metacognition, disciplinary differences, teacher expertise, equity and long-term development of reasoning. Randomized and quasi-experimental study designs, as well as longitudinal designs, shared benchmarks, and clear reporting of model versions will be crucial.¹²

XII. DISCUSSION

The main result is that GenAI is not one educational intervention. Its effects are dependent on the type of model required, what the learner is expected to learn, the information upon which judgment is made, and the retention mechanism of human judgment. This is why you can read in the literature that performance has improved while at the same time seeing very real concerns of dependency or integrity. Any tool that is useful as a formative tutor is not necessarily suitable as an autonomous grader; a model that gets an expert to research quickly can deceive a novice who can't assess the output.

There's also mounting evidence that suggests two simplistic institutional responses are being called into question. The first is the notion of technological solutionism: that is, if you make it better, you'll make the education better. The second is blanket prohibition—the idea that if a technology is widely available, it should be excluded from education and its validity preserved. It is more defensible to redesign learning and assessment to clearly communicate explicit outcomes, and document evidence of performance and learning through the use of AI assistance.

The teacher's role is then made even more significant than before. Educators set epistemic norms of a content area, choose authoritative sources, craft activities that call for reasoning, identify misconceptions, and assess the appropriateness of automation. AI literacy should encompass more than just prompt techniques, however, it should also include information on source evaluation, uncertainty, limitations of models, governance of data, disclosure, and thinking about when AI should not be used.

XIII. RECOMMENDATIONS

Develop institution-wide principles but define acceptable AI use at course and assessment level.

Prioritize low-stakes formative applications before high-stakes automated decisions.

¹⁰ Retrieval-augmented generation for educational application: A systematic survey. (2025). Computers and Education: Artificial Intelligence, 8, 100417. <https://doi.org/10.1016/j.caeai.2025.100417>

¹¹ Large language models for education: A survey and outlook. (2025/2026). IEEE Signal Processing Magazine, 42(6), 51–63. <https://doi.org/10.1109/MSP.2025.3594309>

¹² Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review. (2024). Information, 15(11), 676. <https://doi.org/10.3390/info15110676>

Use trusted-source grounding or RAG where factual reliability and currency matter.
Teach AI literacy as verification, disclosure, privacy awareness, and critical judgment—not merely prompting.
Redesign assessment to capture process, oral explanation, iteration, contextual application, and independent competence.
Protect sensitive data through approved tools, data minimization, retention rules, and procurement review.
Evaluate model performance across disciplines, languages, accessibility needs, and student groups.
Maintain meaningful human oversight and a clear escalation path for consequential educational decisions.

XIV. CONCLUSION

Generative AI and LLMs are becoming a lasting part of the higher-education tech world. There is some evidence of real gains in tutoring, feedback, access, research support and productivity in institutions, yet there is no reason for uncritical automation. Material risks include hallucination, over-reliance, integrity, privacy, bias and unequal access. The most productive direction is human centred augmentation: where AI is used to enhance access to explanation and feedback, anchored in trusted knowledge, checked, reimaged assessment, safeguarded data, and maintained educator responsibility. Advancements should be assessed not just on model fluency and adoption, but on the ability of AI environments to create lasting learning, equitable opportunities, reliable knowledge processes, and enhanced human capacity.

REFERENCES

1. Abdallah, N., Katmah, R., Khalaf, K., & Jelinek, H. F. (2025). Systematic review of ChatGPT in higher education: Navigating impact on learning, wellbeing, and collaboration. *Social Sciences & Humanities Open*, 12, 101866. <https://doi.org/10.1016/j.ssaho.2025.101866>
2. Baig, M. I., & Yadegaridehkordi, E. (2024). ChatGPT in the higher education: A systematic literature review and research challenges. *International Journal of Educational Research*, 127, 102411. <https://doi.org/10.1016/j.ijer.2024.102411>
3. Belkina, M., Daniel, S., Nikolic, S., Haque, R., Lyden, S., Neal, P., Grundy, S., & Hassan, G. M. (2025). Implementing generative AI (GenAI) in higher education: A systematic review of case studies. *Computers and Education: Artificial Intelligence*, 100407. <https://doi.org/10.1016/j.caeai.2025.100407>
4. Guizani, S., Mazhar, T., Shahzad, T., Ahmad, W., Bibi, A., et al. (2025). A systematic literature review to implement large language model in higher education: Issues and solutions. *Discover Education*, 4, 35. <https://doi.org/10.1007/s44217-025-00424-7>
5. Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO.
6. Qian, Y. (2025). Pedagogical applications of generative AI in higher education: A systematic review of the field. *TechTrends*, 69, 1105–1120. <https://doi.org/10.1007/s11528-025-01100-1>
7. Shahzad, T., Mazhar, T., Tariq, M. U., Ahmad, W., Ouahada, K., et al. (2025). A comprehensive review of large language models: Issues and solutions in learning environments. *Discover Sustainability*, 6, 27. <https://doi.org/10.1007/s43621-025-00815-8>
8. UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.
9. Large language models in education: A systematic review of empirical applications, benefits, and challenges. (2026). *Computers and Education: Artificial Intelligence*, 10, 100529. <https://doi.org/10.1016/j.caeai.2025.100529>
10. Retrieval-augmented generation for educational application: A systematic survey. (2025). *Computers and Education: Artificial Intelligence*, 8, 100417. <https://doi.org/10.1016/j.caeai.2025.100417>
11. Large language models for education: A survey and outlook. (2025/2026). *IEEE Signal Processing Magazine*, 42(6), 51–63. <https://doi.org/10.1109/MSP.2025.3594309>
12. Generative AI and higher education: Trends, challenges, and future directions from a systematic literature review. (2024). *Information*, 15(11), 676. <https://doi.org/10.3390/info15110676>