

Use of Large Language Model Tools in Academic Coursework: Opportunities and Challenges in Higher Education

Prasant Kumar¹, Dr Sangeeta Chauhan², Dr Rana Pratap Yadav³

Research Scholar¹, Assistant Professor² & Associate Professor³

¹²³Department of Geography, VSSD College Kanpur U.P. India 208002

Abstract

The rapid development of large language model tools has altered the way university students approach academic coursework. These tools can explain difficult concepts, generate and revise text, provide formative feedback, support brainstorming, assist with programming, translate academic material, and facilitate personalised interaction with learning content. Their growing use creates important opportunities for higher education, particularly where students require immediate feedback or where instructors work with large classes and limited time. At the same time, large language models raise substantial concerns regarding academic integrity, authorship, factual accuracy, fabricated references, over-reliance, privacy, bias, unequal access, and the validity of conventional assessment practices. This paper critically examines the opportunities and challenges associated with the use of large language model tools in academic coursework. It adopts an integrative analytical review based primarily on recent peer-reviewed empirical studies, systematic reviews, and international guidance published since the widespread availability of conversational generative artificial intelligence. The available evidence indicates that carefully structured use can improve academic performance, writing, formative feedback, engagement, and some forms of problem solving. However, educational benefits depend strongly on task design, student artificial-intelligence literacy, verification practices, and continued human judgement. Unrestricted substitution of machine-generated work for students' intellectual effort can weaken independent reasoning and undermine the validity of assessment. The paper argues that higher education should move beyond a simple choice between prohibition and unrestricted adoption. Instead, institutions should integrate large language model tools through transparent rules, process-based assessment, disclosure of material use, source verification, oral or reflective components, and explicit instruction in critical artificial-intelligence literacy. The central educational challenge is therefore not whether students will encounter these tools, but how their use can support rather than replace the intellectual processes that academic coursework is intended to develop.

Article Particulars
Article Title: Use of Large Language Model Tools in Academic Coursework: Opportunities and Challenges in Higher Education
Article Type: Research Article
Volume / Issue: Volume 1 Issue 1
Issue Period: January - March 2026
Publication Date: 12/02/2026
Page Range: 12-25
Licence: CC BY 4.0
Corresponding Author: Prasant Kumar
Email: prasant9456@gmail.com
Authors: Prasant Kumar, Dr Sangeeta Chauhan, Dr Rana Pratap Yadav

Affiliations: VSSD College Kanpur (C.S.J.M University Kanpur)**Peer Review:** Double Blind Peer Review

Keywords: large language models, generative artificial intelligence, academic coursework, higher education, academic integrity, assessment, student learning, artificial-intelligence literacy

1. Introduction

The emergence of large language model tools has changed the technological environment within which university students study, write, solve problems, and complete assessed work. Unlike earlier digital tools that primarily searched, stored, retrieved, or corrected information, contemporary generative systems can participate in sustained dialogue and produce apparently coherent responses to complex prompts. A student may ask such a system to explain a theory, compare competing concepts, suggest an essay structure, revise a paragraph, generate computer code, formulate practice questions, provide feedback on an argument, or simplify difficult academic material. As a result, the boundary between a conventional learning resource and an interactive academic assistant has become increasingly difficult to define.

The significance of this change is particularly evident in coursework-based higher education. Essays, reports, case analyses, reflective assignments, programming tasks, presentations, project proposals, and take-home assessments have traditionally been designed on the assumption that the submitted product provides reasonable evidence of a student's own academic performance. Large language model tools challenge this assumption because they can participate in almost every stage of producing a submission. At the same time, their educational potential cannot be reduced to misconduct. International guidance has emphasised the importance of a human-centred approach in which generative artificial intelligence is governed by educational purpose, human agency, privacy, equity, and pedagogical appropriateness rather than treated merely as a technology to be either accepted or prohibited (Miao & Holmes, 2023).

Research on the educational use of conversational generative artificial intelligence has expanded rapidly. Systematic reviews identify applications across tutoring, feedback, writing, assessment, content generation, language support, programming, and research-related activities, while also documenting concerns about reliability, privacy, fairness, transparency, and academic integrity (Baig & Yadegaridehkordi, 2024; Yan et al., 2024). Experimental evidence is also emerging. A recent meta-analysis of 69 experimental studies found overall improvements in academic performance, affective-motivational outcomes, and higher-order thinking tendencies when ChatGPT was incorporated into learning activities, although methodological limitations and uncertainties about long-term effects require cautious interpretation (Deng et al., 2025).

The central issue, therefore, is not whether large language models are inherently beneficial or harmful. Their educational effects depend on the purpose for which they are used, the nature of the task, the competence of the learner, the quality of instructor guidance, and the extent to which students remain intellectually responsible for what they submit. This paper examines that relationship specifically in academic coursework and considers how higher education can

preserve authentic learning while making responsible use of emerging artificial-intelligence capabilities.

2. Conceptualising Large Language Model Tools in Academic Coursework

Large language models are computational systems trained on very large collections of textual or multimodal information and designed to generate responses by modelling patterns in language and other forms of data. Their practical significance in higher education lies less in their technical architecture than in the range of academic functions they can perform through a single interface. Students can use the same conversational system for explanation, idea generation, editing, translation, summarisation, coding assistance, feedback, question generation, and simulation of academic dialogue.

This versatility distinguishes large language model tools from conventional spelling checkers, search engines, and reference databases. A search engine normally directs the user towards external information, whereas a large language model can immediately synthesise an answer. This efficiency is attractive, but it also changes the student's epistemic relationship with information. Fluent language can create an impression of authority even when a response is incomplete, misleading, or factually incorrect. Educational value therefore depends on whether the learner treats generated output as a provisional contribution requiring judgement or as an authoritative answer requiring little further intellectual work.

Academic coursework is especially sensitive to this distinction. Coursework is intended not merely to produce a polished final document but to require students to read, compare, reason, select evidence, interpret, write, revise, and reflect. If a language model strengthens these processes, it can function as a scaffold. If it replaces them, a high-quality final product may cease to be valid evidence that the student has achieved the intended learning outcomes. For this reason, the more useful question is not simply whether artificial intelligence was used, but what intellectual work was delegated, what decisions remained with the student, whether factual and scholarly claims were independently verified, and whether the student can explain and defend the submitted work.

3. Review of Related Literature

Baig and Yadegaridehkordi (2024) reviewed the growing use of ChatGPT in higher education and found applications across student learning, academic writing, programming, teaching, supervision, assignment design, and administrative activity. Their review demonstrates that generative artificial intelligence has already moved beyond a single-purpose writing aid and is becoming a broad academic support technology. At the same time, adoption is influenced by perceived usefulness, effort expectancy, social influence, institutional policy, and available support, indicating that technology use is shaped as much by educational context as by technical capability.

Research on actual learning outcomes presents encouraging but qualified evidence. Deng et al. (2025) synthesised 69 experimental studies and concluded that ChatGPT interventions generally improved academic performance, affective-motivational states, and higher-order thinking tendencies while reducing mental effort. However, the authors also highlighted limitations such as inadequate power analysis, short intervention periods, and uncertainty over whether improved performance represented durable independent learning. This distinction is

crucial for coursework because assisted performance and long-term competence are not necessarily the same outcome.

Academic writing is one of the areas in which empirical benefits have been most clearly observed. Mahapatra (2024) conducted a mixed-method intervention with undergraduate learners of English as an additional language and found that formative use of ChatGPT was associated with significant improvement in academic writing. The educational mechanism was not simply the generation of finished prose; students could receive rapid feedback, revise their drafts, and engage iteratively with writing problems. Nevertheless, research comparing artificial-intelligence and human feedback suggests that machine feedback should not automatically replace peer or teacher judgement. Banihashem et al. (2024) found that peers were more effective than ChatGPT at identifying specific problems in argumentative essays, even though machine-generated feedback was often more extensive and descriptive.

Evidence from complex problem solving also illustrates both opportunity and caution. Urban et al. (2024) found that university students using ChatGPT produced solutions rated higher in quality, elaboration, and originality and experienced lower perceived mental effort. Yet the relationship between tool use and accurate self-evaluation remained uncertain. This finding raises an important metacognitive concern: when assistance makes a task feel easier, students may not necessarily become better judges of what they themselves know or can do independently.

Student perceptions and patterns of use are also uneven. Baek, Tate, and Warschauer (2024), in a survey of 1,001 college students in the United States, found that patterns of generative-artificial-intelligence use varied according to age, major, institution type, language background, and institutional policy. Non-native English-speaking students reported comparatively frequent use for writing, demonstrating the potential of these tools to reduce language-related barriers. The same study, however, found socioeconomic differences in perceptions, indicating that technology capable of expanding access can also reproduce existing inequalities when access, confidence, and digital literacy are unevenly distributed.

Another major strand of literature concerns academic integrity and assessment. Perkins et al. (2024) showed that relatively simple modifications to generated text can substantially reduce the performance of automated detection systems. This evidence weakens the case for relying on artificial-intelligence detectors as definitive proof of misconduct. It also strengthens the argument for assessment redesign, process evidence, and dialogue with students rather than a technological contest between text generation and text detection. Taken together, the literature suggests that higher education requires a balanced model that uses artificial intelligence as a learning support while preserving verifiable human authorship and intellectual accountability.

4. Objectives of the Study

The study has four principal objectives. First, it examines the major ways in which large language model tools are being used in academic coursework in higher education. Second, it analyses the educational opportunities associated with their use, particularly in relation to personalised learning, feedback, academic writing, accessibility, and problem solving. Third, it identifies the principal challenges involving reliability, fabricated information, academic integrity, cognitive dependence, privacy, bias, equity, and assessment validity. Fourth, it

proposes a balanced framework through which higher education institutions can integrate large language model tools without allowing the technology to replace the intellectual processes that coursework is intended to assess.

5. Research Questions

The study is guided by four questions: How can large language model tools contribute to student learning and the completion of academic coursework? What risks arise when such tools become involved in academic writing, research, problem solving, and assessment? How should higher education institutions redesign coursework and academic-integrity practices in response to these technologies? Finally, what principles can enable students and teachers to use large language models while preserving independent reasoning, source verification, authorship, and responsibility for submitted work?

6. Research Methodology

This study adopts a qualitative, analytical, and integrative review design and is based entirely on secondary sources. The evidence base includes peer-reviewed empirical studies, experimental research, systematic and scoping reviews, and international policy guidance concerning large language models, generative artificial intelligence, higher education, assessment, academic integrity, source reliability, privacy, equity, and student learning. Particular emphasis is placed on research published from 2023 onwards because the widespread educational use of conversational large language models is comparatively recent.

Sources were selected for direct relevance to academic coursework and for the strength of the evidence they provided. Experimental and mixed-method studies were used where the paper discusses learning performance, writing, feedback, or problem solving; systematic reviews were used to identify broader patterns; and international guidance was used for governance, privacy, and institutional policy considerations. The analysis did not treat all forms of artificial-intelligence use as equivalent. Instead, it distinguished between assistance that scaffolds student learning and assistance that substitutes for the cognitive activity being assessed.

The present paper is an integrative review rather than a formal systematic review or statistical meta-analysis. No claim is therefore made that every relevant publication has been included. The objective is to synthesise recent, high-relevance evidence and to develop an academically grounded interpretation of the opportunities and challenges that large language model tools create for coursework in higher education.

7. Opportunities for the Use of Large Language Models in Academic Coursework

7.1 Personalised Learning and On-Demand Explanation

One of the strongest educational opportunities offered by large language models is the possibility of continuous, interactive academic support. University students differ substantially in prior knowledge, language ability, confidence, pace of learning, and willingness to ask questions in formal classroom settings. A conversational system can respond repeatedly to the same learner, explain material at different levels of complexity, create examples, compare concepts, and answer follow-up questions without the time limitations associated with

conventional office hours. Used appropriately, this can extend rather than replace formal teaching.

The educational advantage arises most clearly when interaction is dialogic. A student may ask for an explanation, challenge part of the response, request an alternative example, and then compare the generated explanation with lectures, textbooks, or scholarly literature. This process can support active learning because the tool becomes a means of questioning and elaboration rather than a source of final answers. The risk emerges when students accept the first fluent response without verification or reflection.

7.2 Immediate Formative Feedback

Providing timely individual feedback is among the most demanding aspects of university teaching, especially in large classes. Large language models can reduce the time between student performance and feedback by commenting on draft structure, clarity, argumentation, grammar, coherence, or alignment with assessment criteria. Such feedback can be obtained repeatedly, enabling students to revise work before formal submission. Mahapatra's (2024) findings demonstrate that this model can produce measurable benefits in academic writing when students use generated feedback as part of a revision process.

However, artificial-intelligence feedback should not be assumed to be superior to human feedback. Banihashem et al. (2024) showed that peers were better at pinpointing particular weaknesses in argumentative essays, whereas ChatGPT tended to provide broader descriptive feedback. A combined model may therefore be more educationally productive: automated tools can support rapid first-stage review, peers can contribute critical reading and disciplinary perspective, and instructors can concentrate on deeper conceptual and evaluative issues.

7.3 Support for Academic Writing

Large language models can support academic writing at multiple stages without necessarily writing an assignment on behalf of the student. They can help generate possible questions, organise ideas, test alternative structures, improve transitions, identify unclear sentences, explain grammar, and provide feedback on drafts. For students writing in an additional language, such support can reduce linguistic barriers that might otherwise prevent them from expressing disciplinary knowledge effectively. Baek et al. (2024) found that non-native English-speaking students reported comparatively frequent use of generative artificial intelligence for writing, illustrating this potential accessibility benefit.

The crucial boundary is between support and substitution. A student who drafts an argument and asks for feedback remains directly involved in the intellectual work. A student who asks the system to produce the argument, supporting evidence, structure, and final prose may submit a polished document without having developed the reasoning, research, and writing abilities the task was designed to assess. Higher education therefore needs to define acceptable use in relation to the learning outcome rather than in relation to the technology alone.

7.4 Idea Generation and Complex Problem Solving

Academic work often requires students to move beyond recall towards analysis, synthesis, evaluation, and creative problem solving. Large language models can generate alternative approaches, identify possible variables, propose counterarguments, or expose learners to

perspectives they may not initially consider. Urban et al. (2024) provide experimental evidence that students using ChatGPT produced stronger creative problem-solving responses on several measures. This suggests that large language models can broaden the solution space available to learners and reduce unproductive cognitive bottlenecks.

Nevertheless, better assisted performance does not automatically demonstrate greater independent competence. Students must still evaluate whether a generated solution is conceptually appropriate, methodologically sound, and relevant to the problem. Coursework can make this evaluation explicit by requiring learners to compare their own approach with a generated alternative, identify strengths and weaknesses in both, and justify the final position adopted.

7.5 Accessibility and Inclusive Learning

Large language model tools can support students who experience barriers in conventional academic environments by simplifying dense explanations, providing repeated clarification, offering multilingual assistance, and enabling private questioning without fear of embarrassment. These affordances can be valuable for students with uneven educational preparation and for those who require additional language support. They can also provide alternative forms of explanation when conventional teaching materials are not immediately accessible.

Inclusion, however, is not an automatic consequence of the technology. Effective systems may require paid subscriptions, reliable internet access, suitable devices, and knowledge of how to formulate and refine prompts. Students also differ in their capacity to identify weak or misleading answers. The equity problem therefore involves both access and literacy. If sophisticated assistance is available primarily to students with greater financial and digital resources, artificial intelligence may reinforce rather than reduce educational inequality.

7.6 Support for Teachers and Coursework Design

Academic staff can also use large language models to develop examples, formative exercises, case scenarios, preliminary question banks, marking rubrics, and differentiated explanations. Baig and Yadegaridehkordi (2024) identify staff applications in teaching, supervision, assessment, and academic support. The educational value lies not in replacing professional judgement but in reducing routine preparation time and enabling instructors to focus more intensively on disciplinary interpretation, mentoring, interaction, and high-value feedback.

Teachers can additionally use generated material as an object of critique. A machine-generated answer containing plausible but weak reasoning can become a useful classroom resource if students are asked to identify its assumptions, verify its claims, improve its evidence, and rewrite its conclusion. In such cases, the technology becomes part of the learning task rather than an invisible substitute for it.

8. Challenges Associated with Large Language Model Use

8.1 Inaccurate and Fabricated Information

The fluent style of a large language model can conceal factual errors, oversimplifications, or fabricated sources. These systems are designed to generate plausible responses and do not independently guarantee the truth of every claim. In coursework, this creates a serious risk

because students may mistake polished academic language for verified knowledge. Fabricated or incomplete references are particularly problematic: a citation can look convincing enough to enter an assignment even when the underlying source does not exist or does not support the claim being made.

The appropriate educational response is not simply to warn students that artificial intelligence can be wrong, but to teach verification as a routine academic practice. Students should locate and read every source they cite, check publication details through reliable scholarly databases or publisher pages, and compare important factual claims against authoritative evidence. Large language models should therefore be treated as aids to inquiry, not as substitutes for the scholarly record.

8.2 Over-Reliance and Cognitive Offloading

A second challenge concerns the possibility that convenience may gradually reduce independent intellectual effort. If students routinely delegate brainstorming, summarising, argument construction, interpretation, writing, and revision to a machine, they may become more efficient at producing coursework while becoming less practised in the skills the coursework was intended to develop. This problem can be described as inappropriate cognitive offloading: the learner transfers a process to a tool even though practising that process is itself the educational objective.

Not all reduction of cognitive effort is undesirable. Educational technologies have long removed routine burdens so that students can focus on higher-level reasoning. The relevant question is whether the delegated activity is peripheral to the learning outcome or central to it. A calculator may appropriately remove arithmetic burden in an advanced statistics exercise, while being inappropriate in an assessment of basic arithmetic. The same logic should govern large language model use. Artificial intelligence should reduce avoidable friction without removing the intellectual work that students are expected to learn.

8.3 Academic Integrity and Authorship

Academic coursework normally requires a meaningful relationship between the student receiving the grade and the intellectual work represented by the submission. Large language models complicate this relationship because assistance exists on a continuum, ranging from minor proofreading to near-complete generation of an assignment. Traditional categories such as plagiarism do not fully capture this continuum because machine-generated material may not reproduce identifiable passages from an existing source, yet it can still replace substantial student authorship.

Universities therefore require clear and task-specific definitions of acceptable use. Instructions such as 'declare artificial intelligence use' are insufficient if students do not know whether brainstorming, translation, language correction, summarisation, coding support, restructuring, or text generation must be disclosed. Transparent rules should identify what forms of assistance are permitted for each assessment and require concise disclosure where the use materially affects the submitted work.

8.4 Limitations of Automated Detection

Automated artificial-intelligence detection has been proposed as a response to misuse, but the evidence does not support treating detector output as conclusive proof. Perkins et al. (2024) demonstrated that relatively simple modifications can substantially reduce detector performance. Detection systems may also produce false positives, particularly when writing is formulaic, highly edited, or produced by students whose language patterns differ from the detector's assumptions. High-stakes academic-integrity decisions therefore should not rely on detector scores alone.

A more defensible approach combines assessment design, process evidence, professional judgement, and discussion with the student. Draft histories, annotated sources, oral questioning, practical demonstrations, and reflective explanations can provide much stronger evidence about authorship and understanding than a probabilistic detector. This shifts attention from proving the origin of every sentence to evaluating whether the student can demonstrate the learning represented by the submission.

8.5 Assessment Validity

The deepest challenge is not detection but validity. If a student can produce an acceptable essay without independently demonstrating the knowledge or skills that the essay was intended to measure, the assessment itself may no longer provide valid evidence of learning. This remains true even if the use of artificial intelligence cannot be classified neatly as plagiarism. Large language models therefore require educators to reconsider the relationship between the assessment product and the intended learning outcome.

Coursework can respond by placing greater emphasis on the process through which work is developed. Staged submissions, proposal approval, draft review, source annotation, oral defence, reflective commentary, personalised application, and in-class components can make learning more visible. Such redesign does not eliminate the use of artificial intelligence; instead, it ensures that assistance cannot easily replace all evidence of student understanding.

8.6 Privacy and Data Protection

Students may enter unpublished research data, personal information, assessment questions, institutional documents, or material relating to other individuals into external systems without fully understanding how those data are processed. UNESCO guidance identifies privacy and data governance as central requirements for responsible use of generative artificial intelligence in education (Miao & Holmes, 2023). This issue is particularly important in disciplines involving clinical information, interviews, school records, business data, or confidential organisational material.

Higher education institutions should provide explicit rules on what information may and may not be entered into external systems. Sensitive, identifiable, confidential, or unpublished data should not be shared unless an institution has approved the relevant platform and established appropriate data-processing arrangements. Privacy education should form part of artificial-intelligence literacy rather than being treated solely as an information-technology issue.

8.7 Bias, Representation, and Epistemic Narrowing

Large language models learn from large collections of existing material and can reproduce patterns, assumptions, and biases within those data. They may present dominant academic perspectives more readily than marginalised, local, indigenous, or less digitised knowledge. They may also compress genuinely contested scholarly debates into confident summaries that hide disagreement and uncertainty. In coursework, this creates the risk of epistemic narrowing: students may encounter an apparently balanced answer that is actually shaped by uneven representation in the underlying information environment.

Critical use therefore requires students to ask whose perspective is represented, what evidence is missing, whether the issue is contested, and whether the generated answer is appropriate to the geographical, cultural, disciplinary, and historical context of the assignment. This is particularly important in humanities, social sciences, education, law, and other fields where interpretation and context are central to academic quality.

8.8 Unequal Access and Artificial-Intelligence Literacy

The benefits of generative artificial intelligence are unlikely to be distributed equally. Students with strong disciplinary knowledge, digital literacy, and academic confidence can formulate better prompts, detect weak responses, and combine machine assistance with high-quality scholarly sources. Students with weaker academic foundations may be more likely to accept incorrect responses or allow the tool to make decisions they are not yet able to evaluate. Socioeconomic differences in access to paid systems can create further inequality when more capable models or features are not universally available (Baek et al., 2024).

Equity therefore requires more than making a tool available. Institutions need to teach all students how to use generative systems critically, how to verify claims, how to protect data, how to cite and disclose assistance, and how to recognise when tool use is undermining rather than supporting learning. Artificial-intelligence literacy should become part of general academic literacy.

9. Rethinking Academic Coursework in the Age of Large Language Models

The continued use of assessments designed as though generative artificial intelligence did not exist is increasingly difficult to justify. Coursework needs to move from evaluating only the final submitted product towards collecting evidence of learning across the development process. An essay can remain valuable, but instructors can require a research question, preliminary outline, annotated sources, selected drafts, and a short account of how feedback or artificial-intelligence assistance was used. Oral questioning can then test whether the student understands and can defend central claims.

Coursework can also incorporate artificial intelligence directly into the learning objective. Students may be given a generated answer and asked to identify errors, locate unsupported claims, improve the evidence, identify bias, or reconstruct the answer using verified academic sources. Such tasks develop disciplinary judgement and artificial-intelligence literacy simultaneously. They also recognise that graduates are likely to encounter generative tools in professional settings and therefore need to learn how to evaluate them rather than simply avoid them.

Authentic assessment is likely to become more important. Assignments based on local datasets, practical observations, field experience, laboratory work, personal reflection connected to verifiable activity, institution-specific cases, or staged project development reduce the usefulness of generic generated responses. More importantly, they align assessment with the application of knowledge rather than the production of standardised prose.

10. Discussion

The evidence reviewed in this study suggests that large language models should be understood neither as inherently educational nor inherently harmful. Their impact depends on how they interact with the cognitive work demanded by a particular task. When used to explain, question, provide formative feedback, test alternatives, or expose students to additional perspectives, they can expand opportunities for learning. When used to replace reading, reasoning, evidence evaluation, interpretation, or writing, they can weaken the connection between coursework and learning.

This distinction explains why studies can report improved task performance while concerns about dependence remain valid. A student may produce a better answer with artificial-intelligence assistance. That finding demonstrates improved performance in the assisted condition; it does not by itself demonstrate that the student has acquired the same capability independently. Deng et al. (2025) provide encouraging evidence concerning academic performance and higher-order thinking tendencies, but their analysis also reinforces the need for longer-term and more rigorous measures of learning.

A useful conceptual distinction is therefore between performance with a tool and learning through a tool. Higher education has long accepted calculators, statistical software, reference managers, and spelling tools where those technologies do not invalidate the intended learning outcome. Large language models require the same principle, but on a broader scale because they can participate in language production, reasoning, and knowledge presentation simultaneously. The relevant question is whether the tool supports the intended intellectual activity or replaces it.

A practical framework for responsible coursework can be summarised as assist, verify, disclose, and demonstrate. Students may use an approved tool to assist learning where the assessment permits it; they remain responsible for verifying factual claims and scholarly sources; they disclose material assistance transparently; and they demonstrate understanding through drafts, reflection, oral explanation, practical application, or other evidence appropriate to the discipline. This framework preserves human responsibility while acknowledging that artificial-intelligence-assisted work is becoming part of contemporary academic and professional practice.

The role of the teacher consequently becomes more, rather than less, important. Students need guidance about when generated output is plausible but incorrect, when assistance becomes substitution, how to verify evidence, how to document use, and how to retain intellectual ownership of their work. The central pedagogical task is therefore not merely technical training but the cultivation of judgement. Institutions that focus only on prohibition may fail to prepare students for real-world use, while institutions that allow unrestricted use may weaken the integrity of their own assessments. Responsible integration requires a deliberate middle path.

11. Recommendations for Higher Education

Higher education institutions should adopt clear, differentiated, and assessment-specific rules rather than a single statement that artificial intelligence is either allowed or prohibited. Acceptable use should be directly connected to learning outcomes. Proofreading may be appropriate in one assignment but inappropriate in a language-proficiency assessment; coding assistance may be acceptable in a software-design project but restricted where students are being assessed on basic programming competence. This contextual approach is more defensible than technology-wide rules.

Students should receive structured instruction in artificial-intelligence literacy. Such instruction should cover prompt formulation, verification of claims, checking of references, bias, data privacy, authorship, and disclosure. Particular attention should be given to the fact that fluent answers are not equivalent to verified knowledge. Students should learn to use scholarly databases, publisher pages, and primary sources to verify important claims rather than allowing a conversational interface to become their sole information environment.

Assessment design should progressively incorporate evidence of process. Drafts, source annotations, reflective statements, oral questioning, practical demonstrations, and staged submissions can provide stronger evidence of learning than a single polished final document. Where artificial-intelligence use is permitted, a short disclosure can identify the tool and the purpose for which it was used. Such declarations should be simple, consistent, and non-punitive when use complies with the assessment rules.

Universities should not treat automated artificial-intelligence detection as definitive proof of academic misconduct. Detector output, where used, should function only as one source of information within a fair academic process that includes professional judgement and an opportunity for the student to explain the work. Institutional policy should explicitly recognise the limitations of automated detection.

Staff development is equally important. Responsible integration cannot depend entirely on students when instructors themselves vary greatly in familiarity with generative systems. Teachers require opportunities to understand capabilities and limitations, redesign assessments, create clear instructions, evaluate student disclosures, protect sensitive information, and use artificial intelligence productively in their own academic practice. Effective institutional governance therefore requires the coordinated development of student literacy, staff competence, assessment design, data protection, and academic-integrity procedures.

12. Limitations of the Study

The field is developing rapidly, and findings concerning particular systems can become outdated as models, interfaces, and educational practices change. Much of the empirical literature has been published only since 2023, meaning that evidence about long-term effects on independent writing, disciplinary knowledge, memory, critical thinking, and professional competence remains limited. The current evidence base also spans diverse disciplines, countries, student groups, and intervention designs. A positive effect observed in academic writing cannot automatically be generalised to mathematics, medicine, law, engineering, or laboratory-based education.

The educational consequences of guided classroom use may also differ substantially from unstructured personal use. Many published studies involve explicit researcher or instructor guidance, whereas students using artificial intelligence independently may receive little instruction in verification, disclosure, or privacy. The present paper is an integrative analytical review rather than a formal systematic review and does not estimate a pooled effect size. Future studies should therefore include longitudinal research, discipline-specific comparisons, independent post-intervention assessment, and measures that distinguish immediate improvement in coursework performance from durable learning.

13. Conclusion

Large language model tools are becoming an established part of the academic environment in which university students learn and complete coursework. Their educational potential is substantial. They can provide immediate explanation, personalised assistance, formative feedback, writing support, language assistance, and alternative approaches to problem solving. Recent empirical studies increasingly demonstrate that carefully structured use can improve some forms of academic performance and student engagement.

The same technologies, however, create genuine challenges. They can generate incorrect information and fabricated references, obscure the boundaries of authorship, encourage excessive cognitive delegation, expose confidential data, reproduce bias, intensify inequalities, and weaken the validity of conventional take-home assessment. Automated detection cannot provide a complete or sufficiently reliable solution to these problems.

The appropriate response for higher education is therefore neither unrestricted adoption nor indiscriminate prohibition. Academic institutions should redesign coursework around clear learning outcomes, transparent rules, verification requirements, and stronger evidence of the student's learning process. Students should be allowed to use artificial intelligence where it supports the intended learning, but they must remain responsible for verifying information, acknowledging material assistance, protecting sensitive data, and demonstrating their own understanding.

The long-term educational value of large language models will depend less on the sophistication of the technology than on the quality of the practices surrounding it. When artificial intelligence functions as a scaffold for inquiry, feedback, revision, and reflection, it can extend learning. When it replaces the intellectual effort that coursework is intended to cultivate, it can weaken learning despite producing impressive final work. The central task for higher education is therefore to create academic environments in which technological assistance and human intellectual responsibility coexist.

Declarations

Funding

No external funding was received for this review.

Conflict of Interest

The author reports no conflict of interest in relation to this article.

Ethical Considerations

The study is based exclusively on published literature and publicly available policy documents. It did not involve human participants, personal data collection, or animal subjects; therefore, institutional ethical approval was not required.

Data Availability

No new dataset was generated or analysed. The evidence discussed in the article is available in the published sources listed in the references.

References

- Back, C., Tate, T., & Warschauer, M. (2024). "ChatGPT seems too good to be true": College students' use and perceptions of generative AI. *Computers and Education: Artificial Intelligence*, 7, 100294. <https://doi.org/10.1016/j.caeai.2024.100294>
- 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>
- Banihashem, S. K., Taghizadeh Kerman, N., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: Peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, 21, 23. <https://doi.org/10.1186/s41239-024-00455-4>
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
- Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learning Environments*, 11, 9. <https://doi.org/10.1186/s40561-024-00295-9>
- Miao, F., & Holmes, W. (2023). Guidance for generative AI in education and research. UNESCO.
- Perkins, M., Roe, J., Vu, B. H., Postma, D., Hickerson, D., McGaughran, J., & Khuat, H. Q. (2024). Simple techniques to bypass GenAI text detectors: Implications for inclusive education. *International Journal of Educational Technology in Higher Education*, 21, 53. <https://doi.org/10.1186/s41239-024-00487-w>
- Urban, M., Děchtěrenko, F., Lukavský, J., Hrabalová, V., Svacha, F., Brom, C., & Urban, K. (2024). ChatGPT improves creative problem-solving performance in university students: An experimental study. *Computers & Education*, 215, 105031. <https://doi.org/10.1016/j.compedu.2024.105031>
- Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90–112. <https://doi.org/10.1111/bjet.13370>