International Journal of Learning, Teaching and Educational Research
Vol. 25, No. 8, pp. 909-945, August 2026
https://doi.org/10.26803/ijlter.25.8.36
Received Jun 23, 2026; Revised Aug 4, 2026; Accepted Aug 15, 2026

Mapping of Artificial Intelligence Applications and their Impacts on Academic Performance in Higher Education (2020–2025): A Systematic and Bibliometric Literature Review

Mohammed Al-Battashi*
Department of Islamic History, Civilisation and Education,
Academy of Islamic Studies, Universiti of Malaya,
Sultan Qaboos Naval Academy, Royal Navy of Oman
Sultanate of Oman

Mohamad A. B. M. Adnan
Department of Islamic History, Civilisation and Education,
Academy of Islamic Studies, Universiti of Malaya,
Malaysia

Hassan Abuhassna
UCSI University, Malaysia
Malaysia

Khalid AlFajrani
Faculty of Education, Universiti of Malaya
Malaysia

Citation:
Al-Battashi, M., Adnan, M, A, B, M., Abuhassna, H., & AlFajrani, K., (2026). Mapping of Artificial Intelligence Applications and their Impacts on Academic Performance in Higher Education (2020–2025): A Systematic and Bibliometric Literature Review. *International Journal of Learning, Teaching and Educational Research*, *25*(8), 909–945. https://doi.org/10.26803/ijlter.25.8.36

Abstract. Artificial intelligence is increasingly used in higher education, yet evidence concerning its relationship with academic performance remains fragmented across adoption studies, predictive models and evaluations of learning outcomes. This study maps and synthesises Scopus-indexed research published between 2020 and 2025. A bibliometric analysis of 84 records was conducted using VOSviewer and Microsoft Excel, while 29 eligible studies were examined through a PRISMA-guided systematic review and structured content analysis. The review analysed publication patterns, influential contributors, AI application types, educational purposes, research designs, outcome measures and theoretical frameworks. The findings show rapid publication growth, strong reliance on quantitative and predictive designs, and increasing attention to generative AI, adaptive learning, and learning analytics. However, evidence of improved academic

*Corresponding author: Mohammed Al-Battashi; *batsh14@hotmail.com*

performance varies according to whether studies measured perceptions, prediction accuracy, correlations or direct learning outcomes. Important gaps remain in the limited use of pedagogical frameworks, including TPACK, and the scarcity of experimental, longitudinal, and geographically diverse research. The study contributes an outcome-centred synthesis that connects bibliometric patterns with the methodological and pedagogical strengths of the available evidence.

Keywords: artificial intelligence; academic performance; higher education; bibliometric analysis; systematic review

1. Introduction

Artificial intelligence (AI) has become increasingly prominent in higher education, where it is used in adaptive learning, learning analytics, intelligent support systems, assessment and generative AI applications. These technologies are associated with several forms of academic support, including personalised feedback, early identification of students at risk, and assistance with teaching and learning activities (Alateyyat & Soltan, 2024; Jaboob et al., 2025 Ouyang et al., 2022; Zawacki-Richter et al., 2019). However, the evidence concerning their relationship with academic performance is not uniform. Existing studies differ substantially in what they measure, ranging from technology adoption, perceived usefulness and engagement to prediction accuracy, correlational relationships and directly assessed learning outcomes. Consequently, positive findings reported across the literature should not be treated as equivalent evidence that AI improves academic performance.

The implementation of AI appears to affect its instructional value. Academic-risk prediction systems may identify students, but accurate prediction does not guarantee effective intervention or increased learning. Generated AI technologies may enhance academic writing, feedback and information access but raise concerns about over-reliance, critical thinking, academic integrity and student work quality. These conflicting results indicate that AI should be considered in relation to instructional design, student characteristics, assessment processes, faculty leadership and institutional governance, not only as a technical tool. We need to distinguish between technological capacity or user acceptance and measurable educational impact.

Middle Eastern higher education, where institutional interest in AI is expanding but empirical research is scarce, needs this. Regional research has focused on adoption, readiness, perceptions, and use patterns rather than on longitudinal or intervention-based academic outcomes (Alateyyat & Soltan, 2024; Jaboob et al., 2025). Previous assessments have covered AI applications, online higher education, and adoption patterns. They have not evaluated academic performance, distinguished perception-, prediction-, correlation- and intervention-based evidence consistently or linked these differences to pedagogical frameworks and the bibliometric structure of the field. This hinders our understanding of where and how AI research is progressing; how strong the evidence is and under what educational contexts favourable effects are documented.

This study maps and synthesises AI application studies on higher education academic achievement. It analyses publishing trends, influential contributors, research fields, methodological designs, AI application kinds, academic-outcome measurements and theoretical and pedagogical frameworks using bibliometric analysis and a comprehensive literature review. The systematic review assesses the substantive and methodological qualities of the research, while the bibliometric component provides a macro-level assessment of the evolution and structure of the field. Integrating these two approaches helps the study to discriminate between evidence types, identify methodological, conceptual and geographical gaps and to identify opportunities for pedagogically grounded and context-sensitive research.

2. Literature Review

2.1 AI Applications and Academic Performance in Higher Education

Higher education AI supports or automates teaching, learning, evaluation, and academic decision-making. AI-based systems forecast, personalise, offer feedback, and promote learning using educational data, unlike conventional educational technologies that deliver material (Ouyang, 2022; Zawacki-Richter et al., 2019). Learning analytics, performance-prediction models, adaptive and intelligent learning systems and generative AI tools are the main applications. Cognitive analytics and learning analytics employ academic records, learner characteristics, and digital activity data to evaluate performance, identify students at risk, and inform institutional decision-making. These applications may enable earlier and more tailored support. Prediction accuracy measures the technical performance of a model, not whether it enhances student progress. Educational effectiveness depends on implementing predictions and assessing learning results.

Adaptive and intelligent learning systems tailor content, feedback, pacing and paths to students. Such apps provide tailored advice, skill development, and academic support, although results vary by study and design (Jaboob et al., 2025; Lakhdar et al., 2024). Generational AI supports academic writing, feedback, information access and interaction, expanding its uses. The data suggest benefits and drawbacks for over-reliance, critical thinking, learning quality and academic integrity.

AI applications affect academic achievement conditionally. Studies evaluate model correctness, technological adoption, engagement, self-reported improvement, grades, and test scores, but these are not comparable indicators of educational efficacy. The value of AI depends on its alignment with learning objectives, instructional design, assessment procedures, learner characteristics, and faculty direction. Technical performance and learning outcomes are the best indicators of the educational assistance technology impact of AI.

2.2 Pedagogical, Ethical, and Institutional Conditions for Effective AI Integration

The educational usefulness of AI in higher education depends on technology competence, pedagogy, assessment, institutional conditions, and human oversight. Technology-focused AI applications focus more on efficiency, automation, and predictive accuracy than on learning objectives, teaching strategies, and assessment design (Zawacki-Richter et al., 2019; Ouyang et al., 2022). For weak alignment, AI may enhance operational procedures without providing meaningful or consistent educational advantages.

Pedagogical guidance is crucial for generative AI. These tools provide feedback and learning help, but unstructured or excessive use may promote cognitive offloading, superficial engagement, and diminished higher-order thinking and independent problem-solving (Altinpulluk, 2025). AI should support the intellectual endeavour of students; therefore, instructors must set clear instructional boundaries, devise successful assignments, and define assessment criteria. Many implementations lack proven pedagogical frameworks, making it harder to explain how AI functions affect learning and academic outcomes.

The acceptance and viability of AI-supported education depend on ethics and governance. When institutional regulations or AI-supported judgements are unclear, students and faculty may lose faith owing to data privacy, algorithmic bias, transparency, accountability and academic integrity concerns (Ouyang et al., 2022; Zawacki-Richter et al., 2019). These issues relate to institutional preparation, which includes digital infrastructure, technical support, strategic planning, leadership commitment and responsible data governance (Alateyyat & Soltan, 2024).

Faculty preparation adds to integration success. In practice, the AI literacy, pedagogical competences, attitudes and professional growth of instructors influence tool selection, guiding, and evaluation (Jaboob et al., 2025). Without proper preparation, AI may remain a technology endeavour rather than a meaningful part of curriculum and training. The literature demonstrates that pedagogical congruence, ethical protections, institutional preparation and human competence shape educational efficacy, not technical complexity.

2.3 Regional Perspectives and the Middle East Research Gap

North America, Europe, and East Asia dominate higher education artificial intelligence research, according to the literature. The Middle East has produced little empirical study on the effects of AI on academic performance. The studies region emphasises technology adoption, institutional readiness, and stakeholder perceptions rather than learning results or pedagogically grounded evaluations (Alateyyat & Soltan, 2024; Zawacki-Richter et al., 2019).

Many regional studies are descriptive or exploratory, incorporating few instructional design or learning theories. This hinders their capacity to explain contextually how AI applications affect teaching and student performance (Ouyang et al., 2022). Middle Eastern contexts are under-represented in global AI-in-education research, creating an opportunity for studies that address regional educational priorities, cultural considerations and institutional conditions while contributing to theoretical and empirical discussions.

2.4 Synthesis of Gaps and Direction for the Present Study

Three interconnected deficiencies limit higher education AI research, according to the literature. Many studies focus on adoption, perceived utility, prediction accuracy or correlational relationships rather than on sustained learning or academic success. AI is viewed typically as a technical or analytical tool without considering instructional design, faculty practice, assessment and its pedagogical effects. Ethics, institutional, and human issues are often explored separately rather than in integrated frameworks, and Middle Eastern higher education has few pedagogically grounded and outcome-focused studies (Alateyyat & Soltan, 2024).

Previous reviews have contributed but covered different parts of the field. Zawacki-Richter et al. (2019) mapped higher education AI uses, including research techniques, disciplinary tendencies, and applications. Ouyang et al. (2022) explored empirical studies of AI in online higher education, whereas Jaboob et al. (2025) examined AI adoption trends, opportunities, and difficulties. These reviews did not treat academic performance simultaneously as the main analytical outcome, distinguishing perception-, prediction-, correlation-, and intervention-based evidence and connecting these differences to pedagogical frameworks and to the bibliometric structure of the field.

The present study addresses this intersection through bibliometric mapping of 84 records and systematic analysis of 29 studies published between 2020 and 2025. It links macro-level patterns in publication output, contributors, geographical distribution, and conceptual networks with micro-level evidence on research designs, AI application types, educational purposes, outcome measures, and theoretical foundations. This combined approach allows the review to assess not only how the field has developed, but also what type of evidence supports claims about academic performance.

Accordingly, the study makes four contributions. First, it provides an updated, outcome-centred map of research on AI and academic performance in higher education. Second, it distinguishes technology adoption and perceived usefulness from predictive accuracy, correlational associations, and directly measured learning outcomes. Third, it integrates bibliometric patterns with systematic content analysis rather than presenting the two approaches as separate strands. Fourth, it examines the use of pedagogical and theoretical frameworks, including TPACK, to identify methodological, conceptual, and geographical gaps and to inform more pedagogically grounded and context-sensitive future research.

2.4.1 The bibliometric phase

1. What are the main bibliometric characteristics of research on artificial intelligence applications and academic performance in higher education in terms of the most influential authors, leading contributing countries, most productive institutions, annual publication trends, influential documents, subject-area distribution and predominant research themes and emerging trends?

2.4.2 The systematic literature review phase

1. What are the prevailing research trends in the scholarly literature addressing artificial intelligence applications and their relationship with academic performance in higher education, in terms of publication growth, temporal distribution and research domains?
2. What research methodologies and methodological designs have been employed predominantly in studies examining artificial intelligence applications and academic performance in higher education?
3. What types of artificial intelligence applications have been most frequently investigated in higher education studies, and for what educational and academic purposes have they been utilised?
4. What theoretical frameworks and pedagogical models underpin prior studies exploring artificial intelligence applications and their impact on academic performance in higher education?
5. What future research directions are proposed in the existing literature regarding the use of artificial intelligence applications to enhance academic performance in higher education?
6. What methodological, conceptual, and geographical research gaps exist in the current studies on artificial intelligence applications and academic performance in higher education?

The following sections of this study discuss the approaches utilised to incorporate bibliometric analysis and systematic literature evaluation. A detailed discussion of the data emphasises the important bibliometric and methodological findings. Finally, the article evaluates these results critically and their implications for integrating AI applications into education and improving academic performance and technology to advance the Sustainable Development Goals. These sections cover the aims of the study in detail and relevance.

3. Methodology

The study used a convergent, parallel review design. Bibliometric mapping and systematic content analysis were conducted on related datasets during the same review process and were integrated at the interpretation stage. The bibliometric component described the intellectual and collaborative structure of the field, whereas the systematic component examined the substantive, methodological, and pedagogical characteristics of eligible studies. This design was selected because neither component alone could address both the macro-level structure and the strength and meaning of the evidence concerning academic performance.

3.1 Study Design

The convergent parallel review design combined bibliometric analysis with a systematic literature review. The bibliometric component mapped the broader structure and development of research on artificial intelligence and academic performance in higher education, including publication trends, geographical and institutional distribution, influential contributors, citation patterns and keyword networks. The systematic-review component examined eligible studies in terms of research designs, AI application types, academic-outcome measures, theoretical and pedagogical frameworks, and reported research gaps. Study selection was reported in accordance with PRISMA 2020.

The two components were conducted concurrently using related datasets derived from the same Scopus search and were integrated at the interpretation stage. Macro-level bibliometric patterns were compared with the detailed findings of the systematic review to explain both the development of the field and the nature and strength of the available evidence. This design was selected because bibliometric analysis alone cannot assess the methodological and educational meaning of the literature, while a systematic review alone cannot provide a broad map of the publication and conceptual structure of the field.

3.2 Data Source and Search Strategy

Scopus was selected as the sole bibliographic database for this review because of its broad multidisciplinary coverage and its provision of standardised citation metadata and export formats that are compatible with bibliometric analysis software, such as VOSviewer. These features made it suitable both for identifying relevant publications and for examining publication patterns, citation relationships, and collaboration networks. However, relying on a single database poses a methodological limitation, as it may not capture relevant studies indexed exclusively in other databases. Accordingly, the findings should be interpreted considering this coverage restriction and future reviews may enhance comprehensiveness by incorporating additional databases, such as Web of Science or ERIC.

The search strategy targeted publications addressing artificial intelligence applications, academic performance or learning outcomes, and higher education, in line with the broader aim of the study of examining how AI technologies influence teaching, learning, and academic outcomes in higher education. The review was limited to studies published between 2020 and 2025 to capture recent developments in the educational use of AI, including learning analytics, adaptive learning systems, machine learning, and generative AI. This period provided a clearly defined and manageable corpus for analysing contemporary research patterns and the reported relationship between AI applications and educational outcomes.

3.2.1 Identification

The Scopus database served as a main source of the data for this review. We accessed Scopus on 15 January 2026. We relied on Scopus, as it is one of the most prestigious databases in academic research. This research was built on the relationship between the three main keywords, Artificial Intelligence Applications, Academic Performance and TPACK; for instance, TITLE-ABS-KEY (Artificial Intelligence Applications AND Academic Performance OR TPACK).

3.2.2 Screening

This phase started with 1 312 documents, as we started refining the results, starting with publication years. Therefore, only publications released between 2020 and 2025 were included. Only three subject areas were included in this review: computer science, social science and arts and humanities. Two main document types were included: articles and conference papers. The search process specified a limited number of precise keywords to ensure that the results were more aligned with our research area. Only publications in English were included.

Finally, articles with open access were included. The query used by the researcher to retrieve data and studies related to this topic is as follows:

TITLE-ABS-KEY(Artificial Intelligence Applications AND Academic Performance OR TPACK) AND PUBYEAR > 2019 AND PUBYEAR < 2026 AND (LIMIT-TO (OA, "all")) AND (LIMIT-TO (EXACTKEYWORD, "Artificial Intelligence") OR LIMIT-TO (EXACTKEYWORD, "Education Computing") OR LIMIT-TO (EXACTKEYWORD, "Academic Performance") OR LIMIT-TO (EXACTKEYWORD, "Teaching") OR LIMIT-TO (EXACTKEYWORD, "Performance") OR LIMIT-TO (EXACTKEYWORD, "Higher Education") OR LIMIT-TO (EXACTKEYWORD, "E-learning") OR LIMIT-TO (EXACTKEYWORD, "Personalized Learning") OR LIMIT-TO (EXACTKEYWORD, "High Educations") OR LIMIT-TO (EXACTKEYWORD, "Artificial Intelligence (ai)") OR LIMIT-TO (EXACTKEYWORD, "Educational Technology") OR LIMIT-TO (EXACTKEYWORD, "Artificial Intelligence Technologies") OR LIMIT-TO (EXACTKEYWORD, "Education") OR LIMIT-TO (EXACTKEYWORD, "Student Performance") OR LIMIT-TO (EXACTKEYWORD, "Ai") OR LIMIT-TO (EXACTKEYWORD, "Artificial Intelligence In Education") OR LIMIT-TO (EXACTKEYWORD, "University Students") OR LIMIT-TO (EXACTKEYWORD, "Artificial Intelligence Tools") OR LIMIT-TO (EXACTKEYWORD, "Ai Technologies") OR LIMIT-TO (EXACTKEYWORD, "Educational Innovation")) AND (LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (SUBJAREA, "COMP") OR LIMIT-TO (SUBJAREA, "SOCI") OR LIMIT-TO (SUBJAREA, "ARTS")) AND (LIMIT-TO (LANGUAGE, "English"))

3.2.3 PRISMA 2020 study selection process

The identification, screening and selection of studies were reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). As illustrated in Figure 1, the initial search of the Scopus database identified 1 312 records. Before screening, predefined database limits relating to publication period, subject area, document type, language, and access status were applied. The records were also checked for duplicates. Following the application of these procedures, 182 records remained for screening and full-text retrieval.

The full text was searched of all 182 potentially relevant records. However, 98 reports could not be retrieved through institutional and publisher access channels because they were not downloadable from the Scopus database. Consequently, 84 reports were available in full text for eligibility assessment, and these formed the cohort used in the bibliometric analysis. These 84 reports were evaluated subsequently according to the predefined inclusion and exclusion criteria. At the full-text stage, 55 reports were excluded, leaving 29 pilot studies for inclusion in the systematic literature review. Figure 1 illustrates the number of records identified, removed prior to screening, searched, not retrieved, assessed for eligibility, excluded with explanation and included in each component of the review. The 84 reports retrieved subsequently were evaluated according to the predefined eligibility criteria described in the section that follows.

3.2.4 Eligibility criteria and full-text assessment

A study was eligible for systematic review if it reported primary empirical research conducted in a higher education context, examined an identifiable AI application, and measured academic performance, learning outcomes, or a clearly defined learning-related outcome that consequently correlates with or is targeted by AI use. Studies were excluded if AI was mentioned only as a prospective application, if the sample was outside of higher education, if the study evaluated AI model capability without considering student-level outcomes, or if academic performance was used solely as an input predictor for an unrelated outcome.

Studies were eligible for inclusion in the systematic literature review if they: (a) reported primary empirical research; (b) were conducted wholly or predominantly in a higher education context; (c) examined an identifiable artificial intelligence application, system, or tool; and (d) reported academic performance, learning outcomes, or a clearly defined learning-related outcome as an outcome, correlate, or consequence of AI use or an AI-supported intervention. Studies reporting perceived learning improvement, engagement, self-regulated learning or behavioural outcomes were retained only when these variables were connected explicitly to academic or learning outcomes and were classified as indirect evidence.

Studies were excluded if AI was mentioned only as a prospective or hypothetical application; the sample was drawn from school education or another non-higher education setting; higher education findings could not be separated from findings relating to other educational levels; the study evaluated only the technical capability of an AI model without examining student-level academic or learning outcomes; or academic performance was used solely as an input variable to predict an outcome unrelated to learning improvement. Review articles, conceptual papers, editorials, and other non-primary studies were also excluded.

During the full-text assessment, each excluded report was assigned one primary reason for exclusion to avoid double counting. Of the 84 reports assessed, 55 were excluded: 24 did not measure academic performance or learning outcomes so were excluded for the reasons presented in Table 1; four were unrelated to artificial intelligence applications; seven were conducted outside higher education; seven were outside the substantive scope of the review; 11 were reviews; and two did not meet the language criteria. The remaining 29 studies were included in the systematic literature review.

Table 1: Inclusion and exclusion criteria

#	Dimension	Inclusion criteria	Exclusion criteria
1	Type of study	Primary empirical research	Reviews, conceptual papers, editorials, and non-empirical publications
2	Educational setting	Higher education institutions or clearly separable higher education samples	School education, vocational settings outside higher education, or inseparable mixed settings
3	AI component	An identifiable AI application, model, system, or tool was examined.	AI was mentioned only theoretically or as a future application.
4	Outcome	Academic performance, learning outcomes, or directly related educational outcomes	No student-level academic or learning-related outcome
5	Relationship to AI	The outcome was examined as an effect, correlate, consequence, or target of AI use.	Academic data were used only as input variables for an unrelated prediction.
6	Evidence type	Direct, predictive, correlational, experimental, or clearly classified indirect evidence	Technical AI benchmarking without learner-level outcomes
7	Publication period	Publications between 2020 and 2025	Articles published before 2020 were excluded, owing to the lack of an abundance of scientific papers related to artificial intelligence applications prior to that.
8	Language	English	Languages other than English
9	Document type	Journal articles and eligible conference papers	Other document types are not covered by the review protocol. Thesis and book chapters were excluded.

3.2.5 Data collection and analysis

For the systematic literature review, a final set of 29 articles underwent qualitative analysis to examine key themes, methodologies and challenges. Specific focus areas included stakeholder roles, equity and inclusion in technological innovation and implementation barriers, providing a comprehensive understanding of the subject. The PRISMA framework was used to ensure the transparency and systematic nature of the review process. Bibliometric tools, such as VOSviewer, were used to create visual maps of research trends, while Microsoft Excel was used for statistical analysis. For the systematic review, a structured content analysis framework was applied to categorise the results and identify patterns, as illustrated in Figure 1.

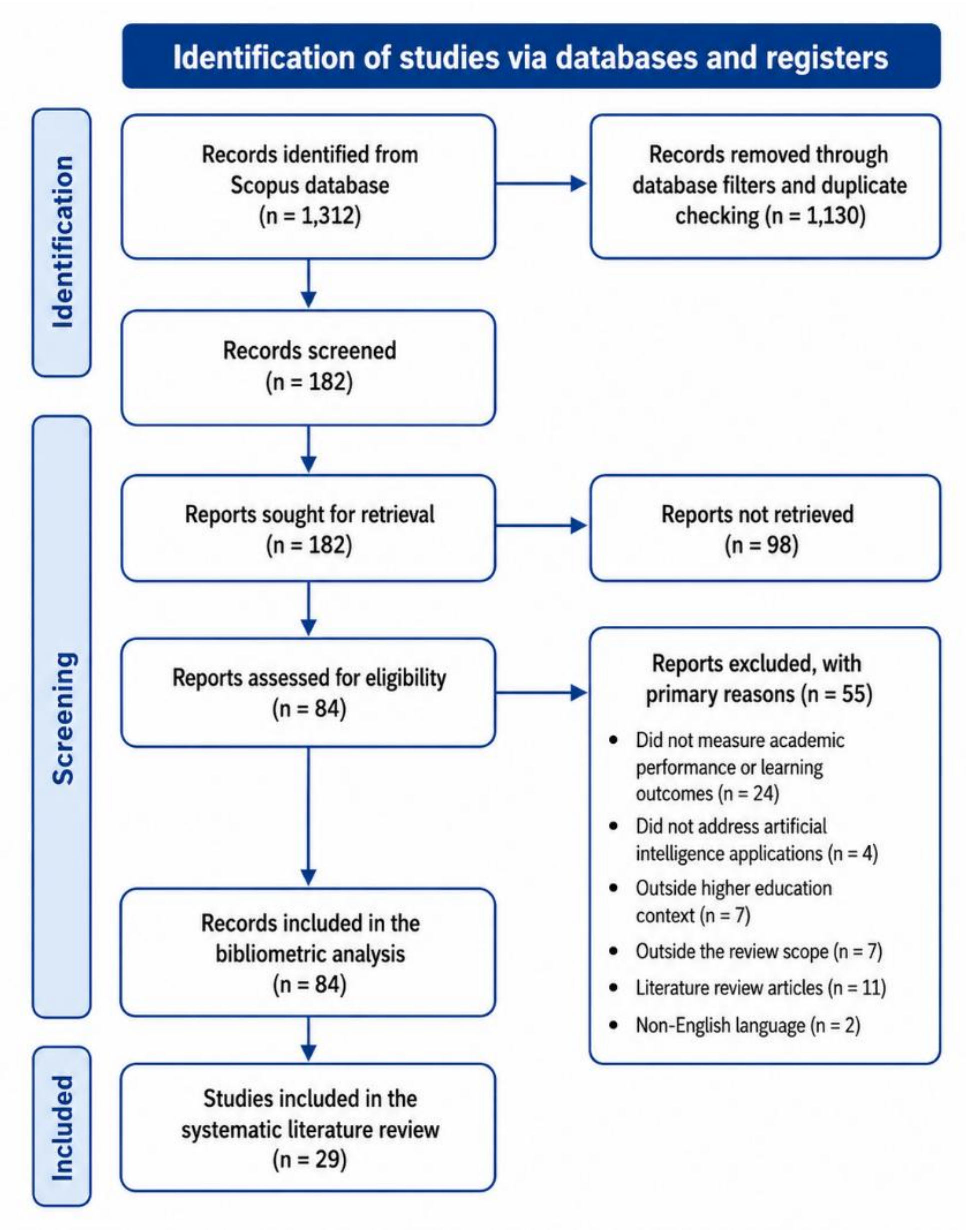


Figure 1: PRISMA framework

4. Results

This review adopts an integrated methodology that combines bibliometric analysis and systematic literature review, with the aim of providing a comprehensive and in-depth understanding of the subject of study.

4.1The Bibliometric Analysis Results

This section uses bibliometric analysis to identify the most influential researchers, the top 20 contributing countries, and leading educational institutions in artificial intelligence applications and their relationship to higher education academic performance. This analysis uses quantitative measures including scientific publication volume, citation impact, and research partnership networks. Our bibliometric presentation covers famous authors, the geographical distribution of scientific output, and major institutions, helping us to comprehend the structure and patterns of research influence in this discipline.

4.1.1 The most influential authors in the field of artificial intelligence applications and academic performance in higher education

Table 2 presents the most productive and influential authors in the field of artificial intelligence applications and academic performance in higher education, based on Scopus author metrics, including total publications, total citations, h-index, and institutional affiliation.

Table 2: The most influential authors in the field

#	Author	SCID*	TP*	TC*	H-Index	Most cited articles	Times cited	Affiliation	Country
1	Villegas-Ch, William Eduardo	57194408086	124	1 552	21	Application of a smart city model to a traditional university campus with a big data architecture: A sustainable smart campus	154	Universidad de las Americas	Ecuador
2	Chamorro-Atalaya, Omar	57729726800	71	300	9	application of the chatbot in university education: A bibliometric analysis of indexed scientific production in SCOPUS, 2013–2023	19	Universidad Nacional Tecnológica De Lima Sur	Peru
3	Al-Ani, Mohammed M.	59340044100	20	244	8	Intelligent system for student performance prediction using machine learning	5	Universiti Kebangsaan Malaysia	Malaysia
4	Abanto-Ramírez, Carlos D.	58979494900	9	14	2	A hybrid AI approach for predicting academic performance in RBE students	0	Universidad Peruana Unión	Peru
5	Abdalkareem, Mai	58908865000	1	1	1	Explainable models for predicting academic pathways for high school students in Saudi Arabia	15	Imam Abdulrahman Bin Faisal University	Saudi Arabia
6	Abdel-Aty, Yasmeen	57211009854	20	68	5	The Bridge2AI-voice application: an initial feasibility study of voice data acquisition through mobile health.	0	Morsani College of Medicine	United States

7	Abdel-Latif, Amr S.	57212264845	4	72	3	System dynamics applications in crisis management: A literature review	6	Alsafa Real Estate Development Inc.	Egypt
8	Abdul Jalil, Rohaya	56579969100	29	152	6	An integrated approach based on artificial intelligence using ANFIS and ANN for multiple-criteria real estate price prediction	6	Universiti Teknologi Malaysia	Malaysia
9	Abdulayeva, Aigerim B.	57239174400	4	13	2	The role of artificial intelligence (AI) in personalised physics education	0	Zhetysu University named after I. Zhansugurov	Kazakhstan
10	Abdulrahman, Ruqayya	36623439600	10	70	4	Fine-tuning OpenAI GPT Chatbot in Western Saudi dialect: a case study of Taibah University	0	Taibah University	Saudi Arabia

SCID = Scopus ID, TP* = Total Publications, TC* = Total Citations*

The results indicate a clear variation in scholarly impact among the leading authors, as reflected by differences in publication volume, citation counts, and h-index values. Villegas-Ch emerges as the most influential author in the dataset, with a notably high number of publications and citations, driven largely by work on smart campus models and big data architectures in university contexts. Other contributors, such as Chamorro-Atalaya and Al-Ani, demonstrate moderate productivity with research focusing on chatbots and machine learning-based student performance prediction, indicating a strong alignment with AI-driven educational analytics.

Table 2 also shows that several authors have produced highly specialised contributions with relatively low citation counts, suggesting that parts of the literature remain emerging or context specific. Geographically, the leading authors are affiliated with institutions across Latin America, the Middle East and Asia, highlighting the global and regionally diverse nature of research on artificial intelligence in higher education. Overall, the distribution of citations and h-index values suggests that, while a small number of authors dominate in terms of impact, the field is characterised by a broad base of contributors exploring varied AI applications related to academic performance.

4.1.2 The highest contributing countries to research on artificial intelligence applications and academic performance in higher education.

Figure 2 illustrates the geographical distribution of published studies on artificial intelligence applications and academic performance in higher education, highlighting the countries with the highest research output.

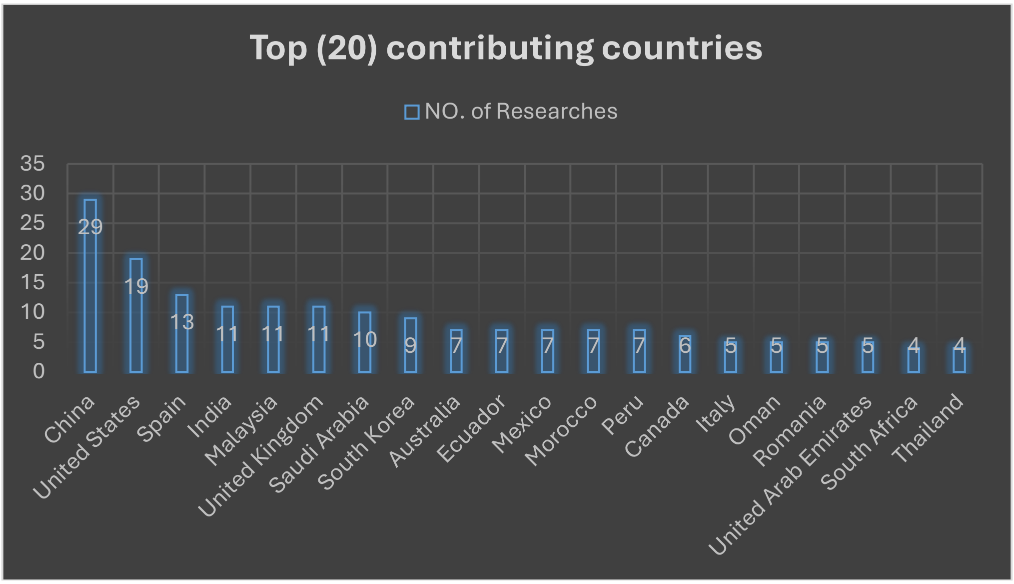


Figure 2: Top 20 contributing countries in the field

Figure 2 reveals a clear concentration of research output in a limited number of countries, with China leading by a significant margin, followed by the United States. This indicates a strong research focus on artificial intelligence and higher education in major research systems. European countries, such as Spain and the United Kingdom, also make significant contributions, reflecting the ongoing interest in AI-led educational research across various regions.

Furthermore, the distribution reveals diverse international participation, with contributions from Asia, the Middle East, Africa, and Latin America. Countries such as Saudi Arabia, Malaysia, Peru, Morocco and Ecuador appear with moderate publication volumes. Several Middle Eastern countries – Saudi Arabia, Oman, and the United Arab Emirates – also feature prominently, indicating a growing interest in AI applications in higher education among these countries in emerging research contexts. Overall, Figure 2 highlights an uneven but globally distributed research landscape, where a few countries dominate the volume of publications, while many others are contributing gradually to the development of this field.

4.1.3 The most productive institutions in research on AI applications and academic performance in higher education

Figure 3 presents the leading academic and research institutions contributing to studies on artificial intelligence applications and academic performance in higher education, based on publication frequency.

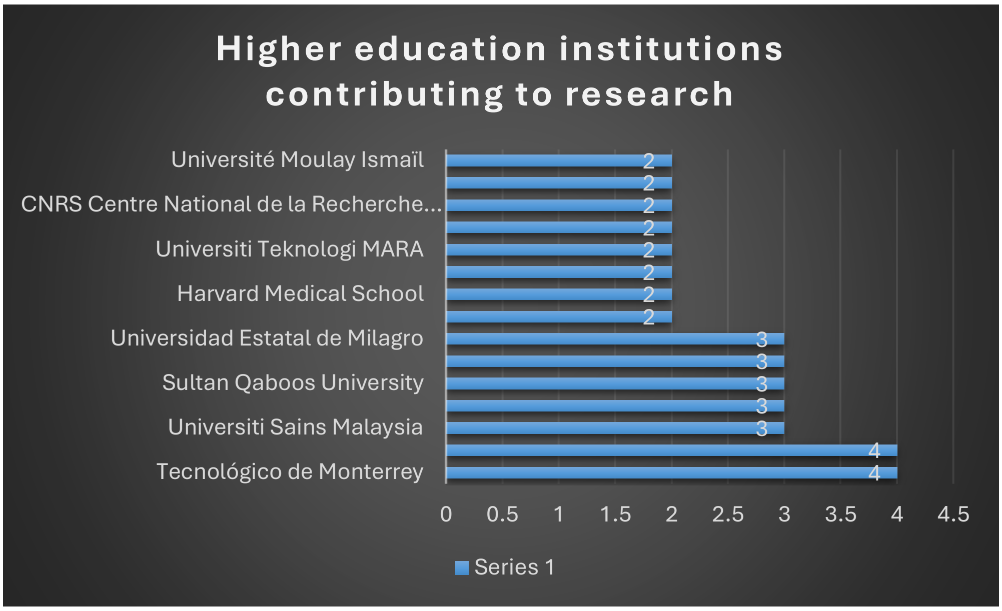


Figure 3: Most productive institutions and research organisations in the field

Figure 3 indicates that research output in this field is distributed across a diverse set of higher education institutions, with no single institution dominating the literature overwhelmingly. Tecnológico de Monterrey and Universidad de las Americas (Ecuador) emerge as the most productive institutions, each contributing the highest number of publications in the dataset. Several universities, including Universiti Kebangsaan Malaysia, Universiti Sains Malaysia, Sultan Qaboos University and Universidad Internacional del Ecuador, also show notable research activity, reflecting sustained institutional engagement with artificial intelligence in higher education contexts.

In addition, the presence of institutions from different regions, including Latin America, Southeast Asia, the Middle East, Europe and North America, highlights the international character of research in this area. Research-intensive organisations and medical schools, such as Harvard Medical School and Weill Cornell Medicine, appear with lower but consistent publication counts, suggesting an interdisciplinary interest in artificial intelligence applications related to education and academic performance. Overall, Figure 3 demonstrates a decentralised institutional landscape, where contributions are spread across multiple universities rather than concentrated in a small number of global research centres.

4.1.4 The most highly published and influential documents by year in the field of artificial intelligence applications and academic performance in higher education

Figure 4 illustrates the temporal distribution of publications on artificial intelligence applications and academic performance in higher education over the period from 2020 to 2025.

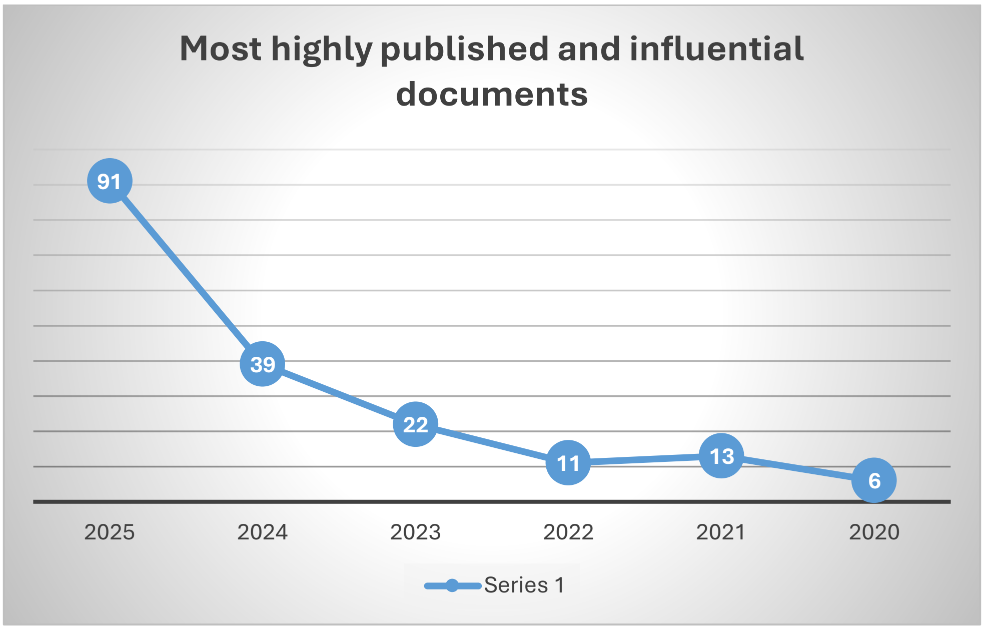


Figure 4: Most highly published and influential documents in this field

Figure 4 reveals a clear upward trend in research output over time, with a relatively low number of publications in the early years, followed by a sharp increase in recent periods. Between 2020 and 2022, the number of studies remains modest, indicating that research on artificial intelligence and academic performance in higher education was still emerging during this phase. A noticeable growth is observed from 2023 onwards, suggesting increased scholarly attention to the topic as artificial intelligence tools become more accessible and adopted more widely in educational settings.

The most pronounced rise occurs in more recent years, with publication counts peaking in 2025. This pattern reflects a rapid acceleration of interest in artificial intelligence applications in higher education, most likely driven by advances in machine learning, learning analytics, and generative AI technologies. Overall, Figure 4 demonstrates that the field has experienced substantial and recent growth, positioning it as a timely and actively expanding area of research.

4.1.5 The top five most published documents by subject area in the field of artificial intelligence applications and academic performance in higher education

Figure 5 depicts the disciplinary distribution of studies on artificial intelligence applications and academic performance in higher education across major academic fields.

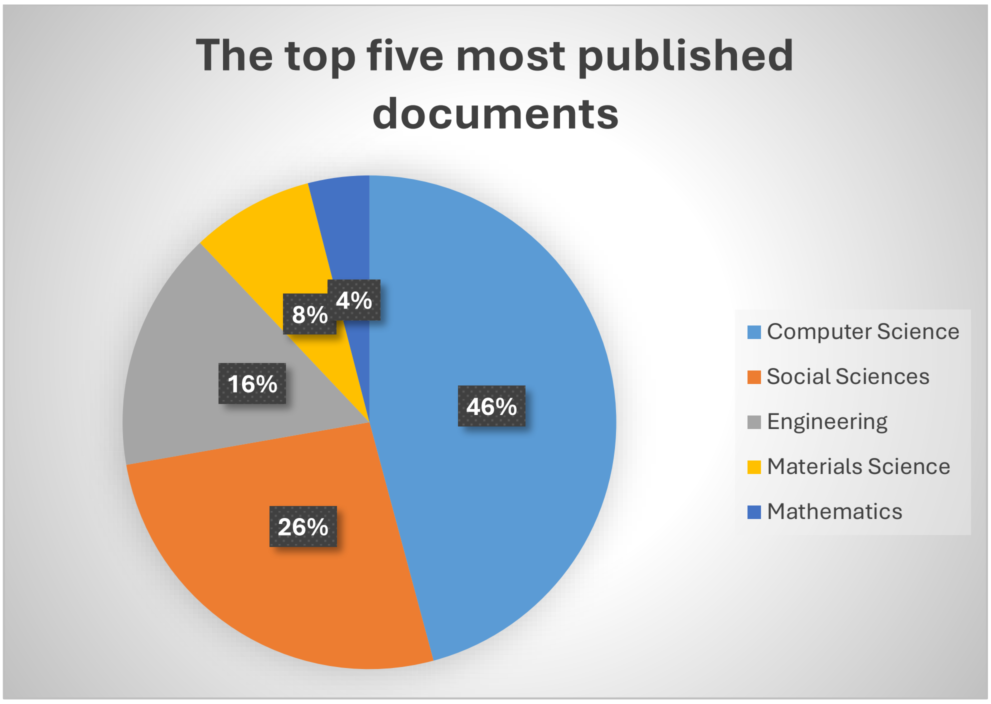


Figure 5: The top five most published documents in this field

Figure 5 shows that research on artificial intelligence in higher education is situated predominantly in the field of computer science, which accounts for nearly half of the published studies. This dominance reflects the technical foundations of artificial intelligence research and the central role of computational methods in developing and evaluating AI-based educational applications. Social sciences represent the second-largest disciplinary contribution, indicating growing interest in the pedagogical, behavioural, and institutional dimensions of AI use in higher education.

Engineering disciplines also contribute a substantial proportion of studies, suggesting an applied orientation towards problem-solving and system development in educational contexts. In contrast, materials science and mathematics account for smaller shares of the literature, reflecting more specialised or indirect engagement with artificial intelligence applications related to academic performance. Overall, Figure 5 highlights a multidisciplinary research landscape, with strong leadership from computer science and increasing contributions from education-related and applied social science fields.

4.1.6 The predominant research themes and emerging trends in studies examining artificial intelligence applications and academic performance in higher education

Figure 6 summarises the most frequently occurring keywords and their total link strength derived from the VOSviewer analysis, providing a quantitative overview of the dominant research concepts in studies on artificial intelligence and academic performance in higher education.

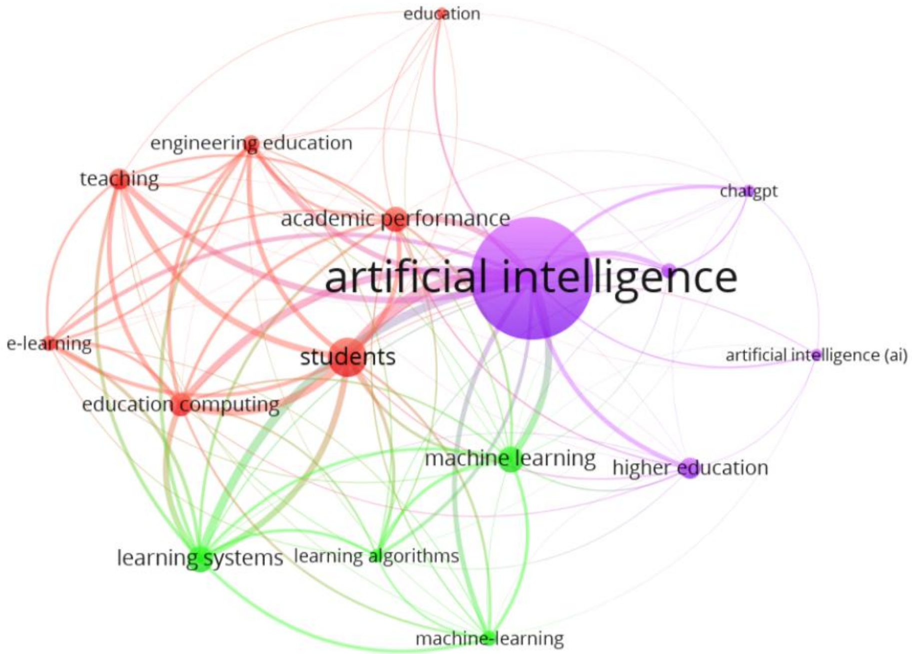


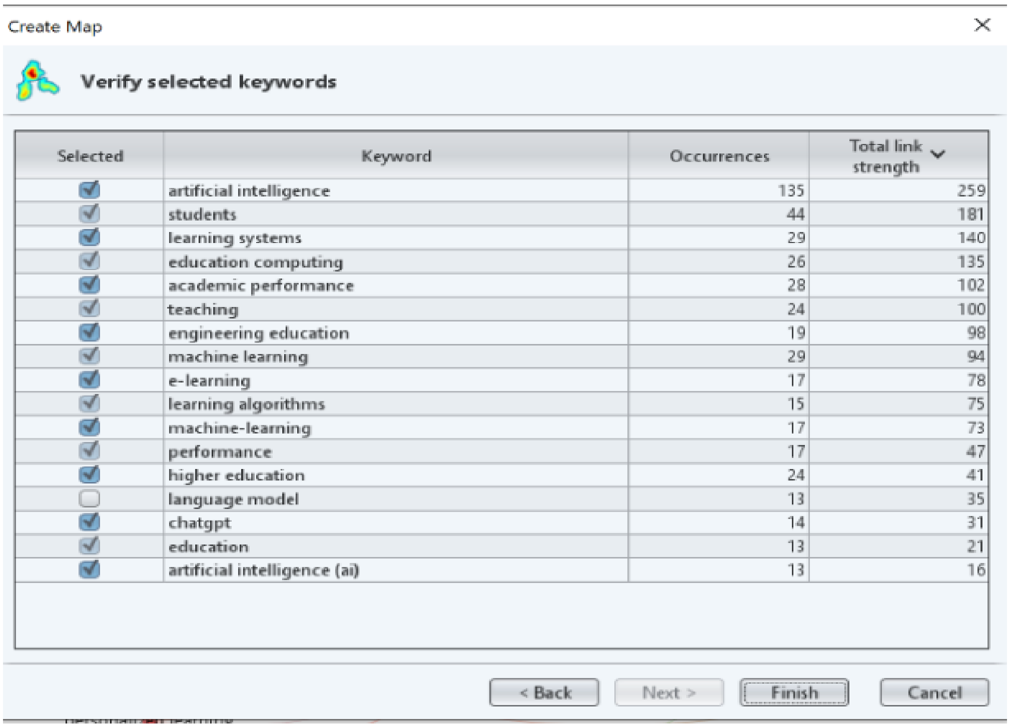
Figure 6: Research themes and trends at the intersection of AI and high performance in higher education institutions

The results indicate that 'artificial intelligence' is the most prominent keyword in the dataset, with 135 occurrences and the highest total link strength (259), confirming its central position across the analysed literature. Keywords related directly to learners and educational processes also show strong representation, including 'students' (44 occurrences; total link strength = 181) and 'learning systems' (29 occurrences; total link strength = 140), highlighting the consistent focus on student-centred AI applications.

Performance-orientated concepts appear prominently, with 'academic performance' recorded 28 times and a total link strength of 102, alongside 'performance' (17 occurrences; total link strength = 47). Teaching-related terms, such as 'teaching' (24 occurrences; total link strength = 100) and 'e-learning' (17 occurrences; total link strength = 78), also reflect the close association between artificial intelligence and instructional practices in higher education contexts.

Figure 7 shows that methodological keywords related to data-driven approaches are also highly visible. 'Machine learning' appears 29 times with a total link strength of 98, while 'learning algorithms' and 'machine learning' record 15 occurrences (total link strength = 75) and 17 occurrences (total link strength = 73), respectively. This distribution indicates a strong reliance on predictive and computational techniques in examining academic performance.

In terms of disciplinary and institutional context, 'engineering education' (19 occurrences; total link strength = 98) and 'education computing' (26 occurrences; total link strength = 135) demonstrate notable connectivity, suggesting that AI applications are explored frequently in technology-orientated educational domains. Keywords referencing the institutional setting explicitly, such as 'higher education' (24 occurrences; total link strength = 41) and 'education' (13 occurrences; total link strength = 21), confirm the higher education focus of the literature further.



Selected	Keyword	Occurrences	Total link strength
✓	artificial intelligence	135	259
✓	students	44	181
✓	learning systems	29	140
✓	education computing	26	135
✓	academic performance	28	102
✓	teaching	24	100
✓	engineering education	19	98
✓	machine learning	29	94
✓	e-learning	17	78
✓	learning algorithms	15	75
✓	machine-learning	17	73
✓	performance	17	47
✓	higher education	24	41
	language model	13	35
✓	chatgpt	14	31
✓	education	13	21
✓	artificial intelligence (ai)	13	16

Figure 7: Methodological keywords related to data-driven approaches

Emerging terms related to generative and language-based AI technologies appear less frequently but remain visible in the network. 'ChatGPT' is recorded 14 times with a total link strength of 31, while 'language model' appears 13 times with a total link strength of 35, indicating a growing yet still developing research emphasis on generative AI tools in relation to academic performance.

4.2 The Systematic Literature Review Analysis Results

This section presents the systematic synthesis of the 29 studies included. The analysis examines publication trends, research designs, AI application types, academic-outcome measures, theoretical frameworks, future research directions, and methodological, conceptual, and geographical gaps. In addition, the review addresses the roles of stakeholders in leading sustainable educational initiatives, highlighting their impact on successful implementation. It also discusses the factors affecting the sustainability and scalability of technological interventions in education, providing a more profound understanding of the elements necessary for long-term success. Finally, this section offers guidance and recommendations for future research and practice to enhance the integration of AI applications in education. This structured approach ensures a comprehensive exploration of the role of AI in particular, and technology in general, in advancing education.

4.2.1 The prevailing research trends in the scholarly literature addressing artificial intelligence applications and their relationship with academic performance in higher education.

Figure 8 summarises the temporal evolution and dominant research domains in studies addressing artificial intelligence applications and academic performance in higher education, as reflected in recent publications from 2020 to 2025.

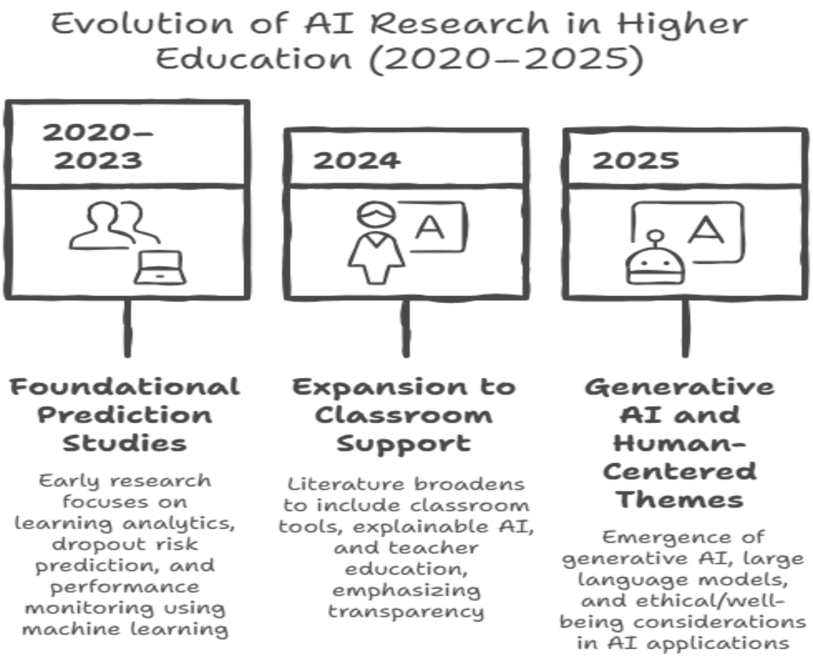


Figure 8: The temporal evolution and dominant research domains in related studies

Table 3 summarises the temporal distribution and principal research domains of the 29 studies included in the systematic review. Publication activity was concentrated strongly in the later years of the review period: of the studies, 72.4% were published in 2025, while 89.7% appeared from 2024–2025. This pattern indicates a recent acceleration in research rather than a steady annual increase. Earlier studies focused mainly on predictive analytics, academic-risk identification, performance classification, and early-warning systems, whereas the recent literature has expanded to include generative AI, academic writing support, AI literacy, personalised learning, classroom assistance, student engagement, and well-being.

In terms of research domains, predictive analytics, educational data mining, early-warning systems, and decision support formed the largest category, representing 41.4% of the reviewed studies. Generative AI, large language models, and AI-supported academic activities constituted the second-largest domain at 31.0%, while adaptive learning, engagement, well-being, technology adoption and sustainability received comparatively less attention. Overall, Table 3 demonstrates a transition from a predominantly technical and predictive orientation towards a broader range of educational applications. However, this thematic expansion does not necessarily reflect stronger evidence of educational effectiveness, as many recent studies measured perceptions, acceptance, behavioural intention or self-reported improvement rather than direct and sustained academic outcomes.

Table 3: Temporal development of the principal research domains

Period	n	%	Dominant research orientation	Main examples
2020–2023	3	10.3%	Academic performance classification, dropout prediction, early-warning systems, and learning analytics	Buranasing (2020), Chavez et al. (2023), Tsai et al. (2020),
2024	5	17.2%	Continued emphasis on predictive modelling and educational data mining, accompanied by explainable AI and the emergence of classroom-support tools	Kwok et al. (2024), Lakhdar et al. (2024), Rico-Juan et al. (2024), Ramírez-Montoya et al. (2024)
2025	21	72.4%	Diversification towards generative AI, AI literacy, adoption, personalised learning, engagement, affective computing, sustainability, academic support, and continued predictive analytics	Studies S01–S23
Total	29	100.0%	–	–

4.2.2 Research methodologies and methodological designs that have been used primarily in studies investigating the applications of artificial intelligence and academic performance in higher education.

Figure 9 shows that most of the reviewed studies use quantitative research methods, with a strong focus on prediction, surveys, and correlation analysis. Many studies rely on machine learning models to predict academic performance, using large institutional datasets and advanced algorithms, such as random forest, support vector machines, gradient boosting, XGBoost and deep learning. These studies usually evaluate their models using standard performance measures like accuracy and error rates, and more recent research also includes methods that help to explain how the models make decisions.

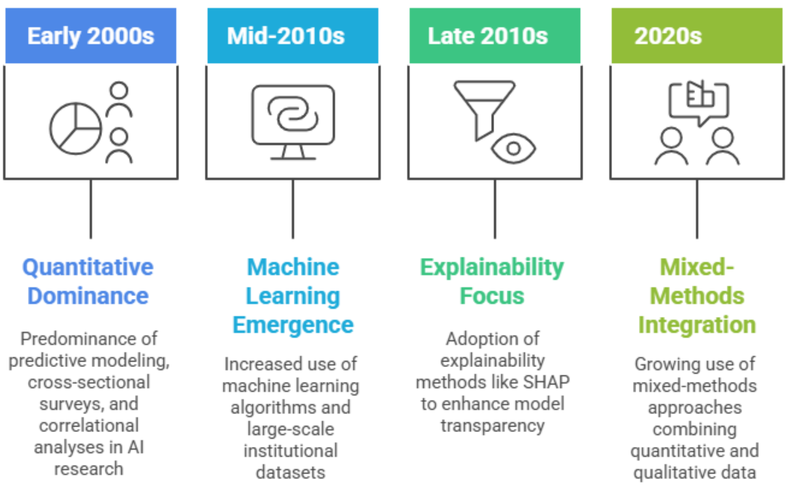


Figure 9: Research methodologies in AI and academic performance studies

Survey-based quantitative studies are also common, especially those examining the adoption of artificial intelligence, user perceptions and behavioural outcomes. These studies often use statistical techniques, such as structural equation modelling and partial least squares analysis, to test the relationships between variables. Most of these studies collect data at one point and focus on testing theoretical models rather than conducting experiments.

Studies that combine quantitative and qualitative methods appear less frequently. These mixed-methods studies often use pre- and post-tests together with interviews or open-ended responses to examine learning outcomes alongside student or teacher experiences. A smaller number of studies follow participants over longer periods, using data collected across multiple years to track changes in academic performance or learning outcomes.

Experimental and quasi-experimental studies are relatively rare and usually involve controlled classroom settings or comparisons between different instructional approaches. Purely qualitative studies, such as those based on interviews or text analysis, make up only a small part of the literature and are used mainly to explore perceptions and experiences related to artificial intelligence in higher education. Overall, Table 3 shows that research in this field relies heavily on quantitative and predictive methods, while long-term, experimental, and in-depth qualitative approaches are used much less often.

4.2.3 Types of artificial intelligence applications and the nature of reported academic outcomes

Figure 10 shows the main types of artificial intelligence applications investigated in the 29 studies included in the systematic review, together with their principal educational purposes. To avoid double counting, each study was assigned to one primary application category according to its central research objective and the main function performed by the AI application.

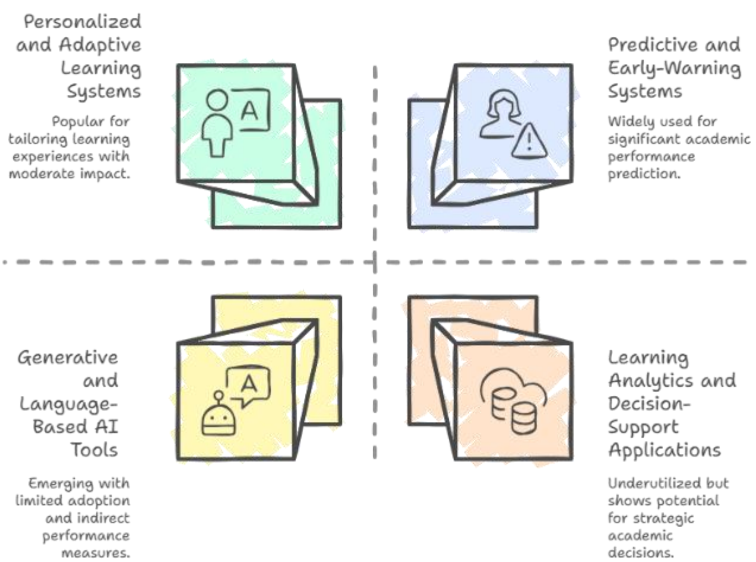


Figure 10: Artificial intelligence applications in higher education

Table 4 shows that predictive analytics, educational data mining, and decision-support applications constituted the largest category, comprising 12 studies (41.4%). These studies employed machine learning, deep learning, learning analytics, effective computing or classification models to predict academic performance, identify students at risk of failure or dropout, support academic advising, and inform institutional decision-making. Common input variables included previous grades, demographic characteristics, learning-management-system activity, attendance, engagement indicators, and student profiles.

Academic performance was operationalised frequently through GPA, course grades, pass–fail status, class ranking or dropout risk. However, in most of these studies, academic performance functioned as a prediction target rather than as an outcome shown to have improved following an AI-supported educational intervention. Measures such as accuracy, precision, recall, F1 score, and prediction error therefore indicate the technical performance of the model rather than direct evidence of improved student learning.

Table 4: Nature of the academic performance evidence reported in the included studies

AI application category	n	%
Predictive analytics, educational data mining, and decision support	12	41.4%
Generative and language-based AI applications	9	31.0%
Personalised, adaptive, and instructional AI systems	5	17.2%
Broader AI adoption and student-support ecosystems	3	10.3%
Total	29	100.0%

Generative and language-based AI applications formed the second-largest category, appearing in nine studies (31.0%). This category included large language models, ChatGPT-like systems, natural language processing applications, AI meeting assistants, and automated programming or academic support tools. These technologies were used for academic writing, summarisation, reflection, literature exploration, programming tasks, automated notetaking, language refinement, assessment support and general study assistance. The reported outcomes included perceived learning improvement, engagement, research efficiency, expected grades, behavioural intention, and willingness to adopt AI. Consequently, positive perceptions or intentions should not be interpreted as direct evidence that generative AI improved formal academic achievement.

Personalised, adaptive, and instructionally orientated AI systems were examined in five studies (17.2%). These applications included adaptive learning systems, learner profiling, intelligent tutoring and support systems, AI-assisted simulation, personalised learning pathways, and AI-supported peer learning. Their main purposes were to adjust content or feedback to individual needs, support engagement, provide targeted academic assistance, and facilitate personalised instruction. Some studies employed experimental or quasi-experimental designs using pre-test/post-test scores, grades or pass rates, while others relied on teacher or student perceptions. The evidence should therefore be interpreted at the individual study level rather than as demonstrating improved academic performance uniformly.

The remaining three studies (10.3%) examined broader AI adoption or student-support ecosystems. These studies addressed institution-level AI services, collections of AI-enabled tools, student satisfaction, sustainability, social integration and well-being. Their outcomes were based predominantly on self-reported academic performance, satisfaction, perceived usefulness, intention to use, or sustainability perceptions. These indicators provide useful evidence concerning the experiences and acceptance of AI by users, but they do not establish changes independently in academic performance measured objectively.

A separate examination of outcome measurement also demonstrates the uneven nature of the evidence. Three studies (10.3%) evaluated academic or learning outcomes following an AI-supported intervention using measures such as pre-test/post-test scores, grades or pass rates. Eleven studies (37.9%) used objective academic records as prediction or association targets but did not test whether the AI application itself produced improvement. Another 12 studies (41.4%) relied primarily on perceptions, engagement, behavioural intention, expected grades, self-reported performance or other indirect indicators. The remaining three studies (10.3%) examined technical model capability or prospective AI use without measuring student-level learning improvement.

The reviewed literature indicates that researchers have investigated AI for performance prediction, personalised support, academic writing, student engagement, institutional decision-making and other academic support activities. However, one should distinguish between the frequency of application studies and the strength of evidence supporting their educational effectiveness. The corpus contains substantial evidence concerning predictive capability, user perceptions, engagement, and intention to use, but there is considerably less direct evidence that AI-supported interventions produce sustained improvements in academic performance measured objectively.

4.2.4 The most prominent theoretical frameworks and educational models used in the included studies

This section examines the theoretical and pedagogical frameworks reported explicitly in the 29 studies included in the systematic review. Table 5 summarises the theoretical and pedagogical frameworks underpinning the selected studies on artificial intelligence applications and academic performance in higher education.

Table 5: The theoretical and educational frameworks upon which the selected studies were based

Category	Number of studies	Percentage
Studies reporting an identifiable framework or theoretical perspective	25	86.2%
Studies relying on implicit orientations without an explicit framework	4	13.8%
Technology acceptance and readiness models	5	17.2%
Technical and analytical frameworks	11	37.9%
Learning, motivational, or instructional perspectives	16	55.2%
Technology–pedagogy integration frameworks	0	0.0%
TPACK	0	0.0%
SAMR	0	0.0%

Overall, 25 studies (86.2%) reported at least one identifiable theoretical framework or perspective, while four studies (13.8%) relied mainly on implicit conceptual orientations. Because some studies used multiple frameworks, the categories were coded independently and were not mutually exclusive; therefore, their percentages do not total 100%.

Technology acceptance and readiness models appeared in five studies (17.2%). These included UTAUT, extended UTAUT, TRAM, TAM2, and TAM/TAM3 extensions (Bećirović et al., 2025; BinJwair, 2025; Campillo-Ferrer et al., 2025; El Hosayny et al., 2025; Krishnan et al., 2025). These models mainly examined behavioural intention, perceived usefulness, trust, readiness, and willingness to adopt AI. Their focus was therefore on explaining AI acceptance rather than on its alignment with pedagogy, disciplinary content and learning activities.

Eleven studies (37.9%) were grounded primarily in technical or analytical perspectives, including learning analytics, educational data mining, predictive analytics, benchmarking, explainable AI, assessment theory and precision education. These studies mainly assessed model accuracy, algorithmic performance, interpretability, academic-risk identification, and the reliability of AI-supported prediction or assessment systems. While discussions often centred on educational applications, the primary focus remained on computational performance rather than on pedagogical integration.

Learning, motivational, or instructional perspectives were identified in 16 studies (55.2%). These included gamification, engagement theory, adaptive and personalised learning, experiential learning, constructivism, Bloom's Taxonomy, cognitive load theory, self-efficacy theory, self-regulated learning, AI literacy, and the Precision Education framework. For example, Yousef et al. (2025) used Bloom's Taxonomy, Kwok et al. (2024) applied cognitive load and self-efficacy theories, and Rico-Juan et al. (2024) combined the model of human performance with self-regulated learning. However, none of the 29 studies operationalised TPACK or SAMR explicitly as a technology–pedagogy integration framework.

The absence of TPACK reflects the dominant orientation of the reviewed literature. Many studies focused on prediction, educational data mining, model evaluation or institutional decision support, while acceptance-based studies concentrated on usefulness, trust, readiness, and intention to use. Relatively few examined subject-specific classroom interventions in which AI was integrated systematically with curriculum content, teaching strategies, and assessment practices in which TPACK would be most applicable.

This absence should not be interpreted as a complete lack of theoretical grounding. Rather, it indicates that the literature has prioritised predictive, adoption-orientated, and technically focused research over pedagogical integration. Consequently, considerably more is known about whether AI can predict performance or whether users are willing to adopt it than about how instructors can integrate AI to produce measurable learning gains. Future studies should therefore apply frameworks that connect technological affordances with pedagogy, disciplinary content, assessment and academic outcomes.

4.2.5 Future research directions suggested in the current literature regarding the use of artificial intelligence applications to enhance academic performance in higher education.

Figure 11 synthesises the future research directions proposed across the selected studies, outlining the methodological, contextual, and conceptual priorities identified for advancing research on artificial intelligence and academic performance in higher education. Figure 11 summarises the main directions for future research suggested by the reviewed studies, focusing on how research on artificial intelligence and academic performance in higher education can be improved and expanded.

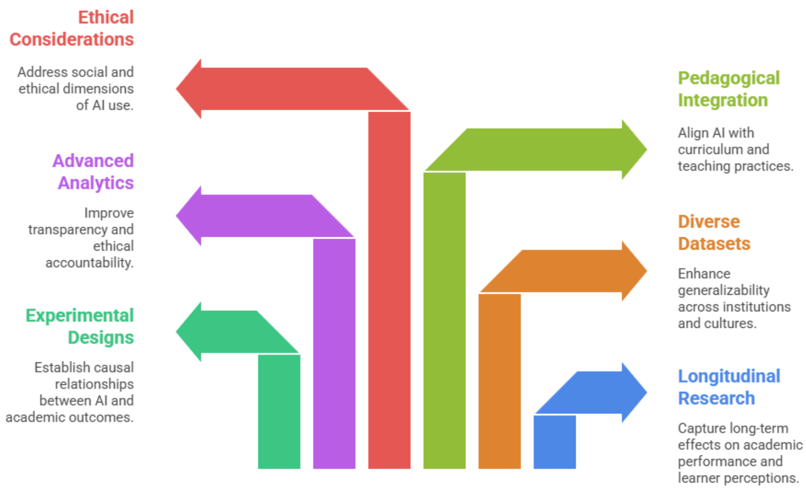


Figure 11: Future research directions for improving academic performance in higher education

Many studies strongly recommend using longer-term (longitudinal) research designs. These designs would allow researchers to understand better the lasting academic, cognitive, and behavioural effects of AI-supported learning. Long-term studies are considered important for tracking changes in academic performance over time, understanding how the views of students evolve, and examining the long-term impact of AI on institutions, rather than relying only on short-term or one-time data.

Another common recommendation is the use of experimental and quasi-experimental research designs. Studies suggest using pre- and post-tests, controlled comparisons between AI-supported and traditional learning environments and intervention-based evaluations in learning management systems. These approaches are needed to determine clearly whether AI applications cause improvements in academic performance, rather than simply showing correlations.

Figure 11 also emphasises the necessity of larger and more diverse datasets, including studies that involve multiple institutions, disciplines and countries. Researchers emphasise that testing AI models across different educational contexts would improve the reliability and general usefulness of research findings, compared to studies limited to a single institution or discipline.

Several studies call for the use of richer data sources and more advanced analytical methods, such as combining academic data with behavioural, emotional and digital activity data. There is also growing interest in using explainable AI methods and fairness checks to make AI systems more transparent, reliable and ethically responsible.

Moreover, educators are increasingly focusing on integrating pedagogy and conducting research at the curriculum level. Studies recommend aligning AI tools with instructional design, curriculum development, and subject-specific teaching practices. This includes integrating AI literacy, ethical use of AI and critical digital skills into academic programmes, as well as examining the role of instructors, teacher training and institutional policies in supporting AI-based learning.

Finally, ethical issues are highlighted as an important future research area. Many studies stress the need to examine topics, such as fairness, trust, transparency, data privacy and governance. These recommendations reflect a growing understanding that future research should address not only the technical performance of AI systems but also their social, ethical and institutional impact in higher education.

4.2.6 Methodological, conceptual, and geographical research gaps.

Figure 12 summarises the methodological, conceptual and geographical gaps identified across the studies included, based on their designs, theoretical foundations, contexts, reported limitations and future research directions.

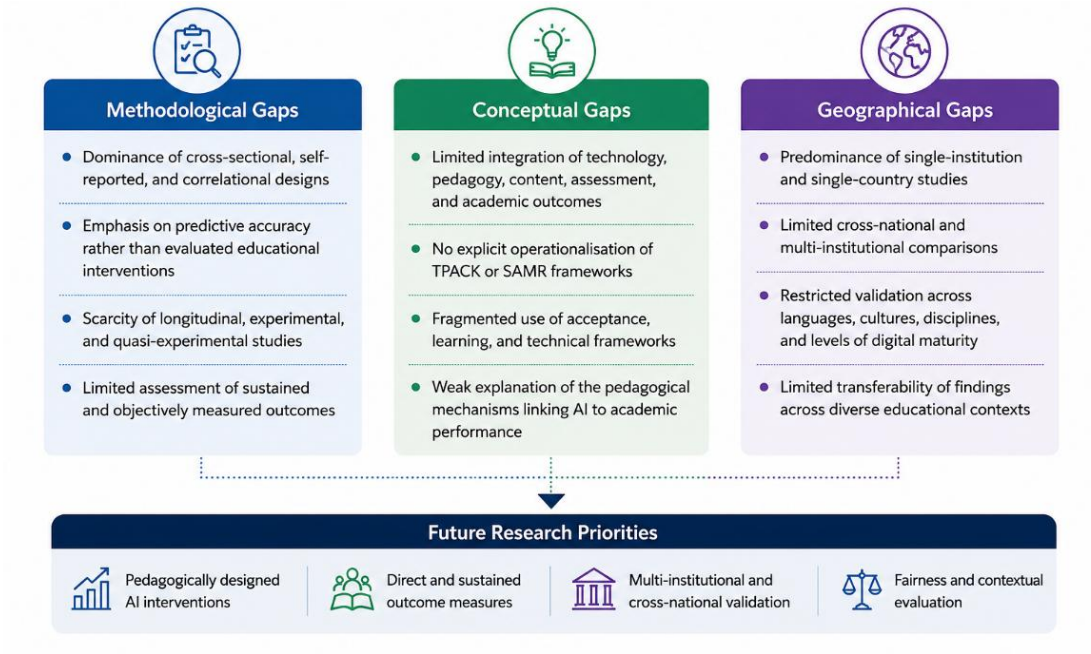


Figure 12: Methodological, conceptual, and geographical gaps identified

Methodologically, literature is dominated by cross-sectional surveys, self-reported measures, correlational analyses and predictive modelling. These approaches are useful for examining AI adoption, perceptions, behavioural intentions and factors associated with academic performance, but they provide limited evidence that AI use produces sustained improvements in learning outcomes. Many predictive studies evaluated models using technical indicators, such as accuracy, precision, recall and F1 score, without testing whether the resulting predictions were translated into effective academic interventions or measurable learning gains. Longitudinal, experimental, quasi-experimental and multi-institutional studies remained comparatively uncommon, while some classroom interventions relied on small samples, short implementation periods or perceived improvement rather than on objective and sustained academic outcomes.

A related gap concerns the distinction between predictive performance and educational effectiveness. Identifying students at risk does not demonstrate that an AI system improves academic performance. Such a claim requires an intervention pathway in which AI-generated information informs academic support, instructional adaptation, or personalised feedback, followed by a direct evaluation of the resulting outcomes. This pathway was rarely examined systematically in the predictive studies reviewed.

Conceptually, the literature showed limited use of integrated frameworks connecting technological capabilities with pedagogy, disciplinary content, assessment and academic outcomes. None of the 29 studies included operationalised TPACK or SAMR explicitly in the design or evaluation of AI-supported instruction. Technology acceptance models, such as TAM, UTAUT, and TRAM, explained perceived usefulness, readiness, trust and intention to use but offered limited insight into how AI should be embedded in specific teaching practices. Likewise, learning analytics, educational data mining, predictive analytics and explainable AI studies focused mainly on data, model performance and interpretability. Although some studies employed learning or instructional perspectives, generally these were applied separately rather than in a comprehensive framework linking AI functionality, instructor decisions, subject content, learning activities and outcome assessment. Consequently, the pedagogical mechanisms through which AI may contribute to academic performance remain insufficiently specified.

Geographically, the studies represented several regions, including Europe, Asia, the Middle East, Africa and Latin America, but most were confined to a single institution, programme, discipline or country. The limited number of cross-national comparisons and studies involving institutions with varying levels of digital maturity hindered the research. This restricts the transferability of findings and makes it difficult to determine whether AI-supported practices and predictive models perform consistently across educational systems, languages, cultures, disciplines and socioeconomic contexts.

5. Discussion

The findings indicate that the relationship between artificial intelligence applications and academic performance is promising but not uniform. The studies reviewed examined diverse applications, including predictive analytics, adaptive and personalised learning, generative AI, natural language processing, intelligent tutoring, affective computing, automated assessment and student-support systems. They also measured different outcomes, ranging from actual grades and pass–fail results to engagement, self-efficacy, perceived improvement, technology acceptance and model accuracy. Therefore, the evidence does not support a general claim that AI improves academic performance consistently. Its educational value depends on the type of application, its pedagogical integration and the outcome measures used.

The clearest evidence of educational improvement came from experimental or quasi-experimental interventions. Abdulayeva et al. (2025) examined an adaptive physics-learning system combining learner profiling, adaptive content, predictive analytics and intelligent problem-solving support. Kasimovskaya et al. (2025) similarly evaluated AI assistants and HoloHuman through HoloLens in medical education, linking simulated ethical scenarios and adaptive feedback with academic performance, self-efficacy and moral behaviour. These findings suggest that AI is more likely to support measurable learning outcomes when embedded in a defined instructional intervention. However, the relatively short duration and context-specific samples of these studies restrict conclusions about long-term effectiveness and transferability across disciplines and institutions.

A substantial proportion of the literature focused on predicting academic performance rather than on improving it directly. Machine learning and learning analytics studies used institutional records, prior grades, attendance, LMS activity and behavioural variables to classify achievement and identify students at risk of failure or dropout. This pattern appeared in Balderrama and Araújo Lima (2024), Buranasing (2020), Chavez et al. (2023), Kustitskaya et al. (2025), Lakhdar et al. (2024), Ramírez-Montoya et al. (2024), Rico-Juan et al. (2024), Tsai et al. (2020), and Yan and Yang (2025). Although several models achieved high predictive accuracy, such results indicate an ability to estimate outcomes, not evidence that the models improved them. Most studies did not evaluate whether predictions were followed by instructional interventions or whether such interventions produced measurable learning gains.

Evidence concerning generative and language-based AI was mainly indirect. Campillo-Ferrer et al. (2025) associated generative AI with content creation, personalisation, collaboration, critical thinking and study support but assessed improvement primarily through student perceptions. BinJwair (2025) and Krishnan et al. (2025) examined behavioural intention, performance expectancy, trust, adoption and sustainability rather than objective achievements. El Hosayny et al. (2025) reported that doctoral students used ChatGPT, Grammarly, Copilot and Gemini for literature exploration, summarisation, writing, language refinement and conceptual clarification, while Yousef et al. (2025) found perceived academic benefits among medical students. Nevertheless, these outcomes reflected perceived efficiency, writing quality, research support or expected improvement rather than learning gains that were measured directly and sustained.

The findings also show that AI effects vary across disciplinary and instructional contexts. Anwar et al. (2025) identified different relationships among CourseMIRROR engagement, academic engagement and performance across engineering and physics courses. Kwok et al. (2024) examined Otter.ai for transcription, summarisation and note-taking but measured performance through expected rather than actual grades. El Bahri et al. (2025) linked the emotional states of students with attendance and grades, although the findings remained correlational. Similarly, Wilson-Trollip (2025) found that an AI-supported peer-learning platform enhanced engagement and support, but improvements in grades and pass rates were largely insignificant statistically. These results indicate that engagement, satisfaction, confidence and well-being may support academic success, but they should not be treated as equivalent to achievement measured objectively.

The theoretical analysis revealed diversity but limited integration. Some studies employed UTAUT, extended UTAUT, TRAM, TAM2, and TAM/TAM3 to explain acceptance, usefulness, trust and readiness. Others used engagement theory, adaptive learning, cognitive load, self-efficacy, Bloom's Taxonomy, self-regulated learning, the Model of Human Performance and Precision Education. However, none operationalised TPACK or SAMR explicitly to connect AI functionality with pedagogy, disciplinary content and assessment. This does not imply an absence of theory; rather, it reflects the dominance of predictive and adoption-orientated studies over research examining pedagogical integration.

Methodologically, the field was dominated by predictive, cross-sectional, correlational and perception-based designs, while experimental, longitudinal, and mixed-methods interventions remained limited. Many studies were also conducted in single institutions, disciplines or national contexts, restricting generalisability. Overall, AI appears to influence academic performance through a conditional pathway: it provides predictions, feedback, personalisation or content support. Educators translate these outputs into pedagogical actions; and these actions may influence engagement, self-efficacy, feedback quality and learning efficiency before affecting achievement. Future research should therefore connect AI-supported interventions with objective, sustained outcomes and examine faculty roles, disciplinary variation, ethics, fairness, privacy and institutional governance through integrative pedagogical frameworks such as TPACK.

6. Implications for Research, Policy and Higher Education Practices

The findings of this review point to implications for future research, institutional practices and AI governance in higher education. These implications arise directly from the methodological, conceptual and geographical gaps identified in the included studies.

6.1 Research Implications

Future studies should employ more longitudinal, experimental and quasi-experimental designs to assess the academic effects of AI-assisted interventions over time. Researchers should distinguish clearly between predictive model performance, technology uptake, perceived learning improvement and academic outcomes measured objectively. Predictive accuracy should not be interpreted as educational effectiveness unless AI-derived insights lead to a specific intervention whose impacts are assessed through grades, assessment results, skills development, student retention or other direct learning indicators.

Furthermore, multi-institutional and cross-national studies are needed to examine the reproducibility of findings across different disciplines, languages, cultures and levels of digital maturity. Research should address the nature and duration of the intervention, the role of the instructor, the educational objective of the AI application and the outcome measures used. Educational frameworks, such as the TPACK framework or AI-specific plugins, can help to clarify how technological capabilities interact with subject-matter content, teaching strategies, assessment and learner characteristics.

6.2 Implications for Higher Education Practices

Universities should complement technical training with professional development that focuses on pedagogical integration. Faculty members should be supported in aligning the use of AI with learning objectives, designing appropriate AI-enabled activities, evaluating their outcomes and fostering critical thinking and self-directed learning. AI-based early-warning systems should function as decision-support tools, not as independent decision-makers.

6.3 Implications for Policy and Governance

Institutional policies should address academic integrity, privacy, data security, transparency, accountability and the responsible use of data. The acceptable uses of generative AI, the responsibilities of students and teachers, the types of data and their collection objectives and the procedures for reviewing or challenging automated recommendations should all be defined clearly. AI systems should be piloted and evaluated for their educational effectiveness, technical reliability and fairness across different student groups, disciplines and relevant study modes. Traditional performance evaluations must be accompanied by fairness assessments and human oversight. Scaling up should only be permitted once the application has demonstrated its educational value, complies with privacy and integrity requirements and is subject to ongoing monitoring.

7. Limitations

Several limitations should be considered when interpreting the findings of this review. First, the search was restricted to the Scopus database. Although Scopus provides broad multidisciplinary coverage and structured bibliographic metadata, relevant studies indexed exclusively in other databases may have been omitted. This single-database approach may therefore have influenced the geographical, disciplinary and publication patterns identified in both the bibliometric and systematic analyses.

Second, if open-access status was retained as an eligibility restriction, the review may over-represent studies that were accessible freely while excluding relevant subscription-based publications. Such a restriction could affect the composition of the evidence base and may introduce accessibility bias, particularly across countries and institutions with different publishing and funding practices. The open-access restriction should therefore be reported explicitly only if it was applied during study selection.

Third, bibliometric indicators, including citation counts and collaboration measures, are dependent on the coverage and update cycle of the selected database and on the date of data extraction. Citation-based results may favour older publications because they have had more time to accumulate citations, while recently published studies may appear less influential despite their potential relevance. Accordingly, citation impact should be interpreted as a time-sensitive indicator of scholarly visibility rather than as a direct measure of methodological quality or educational effectiveness.

Fourth, the studies included employed heterogeneous definitions and measures of academic performance and learning outcomes. These measures included grades, GPA, pass–fail status, prediction accuracy, engagement, behavioural intention, self-reported improvement, writing quality and other indirect indicators. This heterogeneity limits direct comparison across studies and prevents the findings from being interpreted as a uniform estimate of the effect of AI on academic performance.

Fifth, restricting the review to English-language publications may have resulted in language bias and the under-representation of research published in Arabic and other languages. In addition, publication bias may be present because studies reporting positive or statistically significant findings are more likely to be published and indexed than studies reporting null, negative or inconclusive outcomes. Consequently, the literature reviewed may present a more favourable account of AI applications than the full body of conducted research.

Finally, the systematic review included a relatively limited and methodologically diverse corpus. The predominance of cross-sectional, predictive and self-reported studies, together with the limited number of longitudinal and experimentally evaluated interventions, restricts the strength of causal conclusions that can be drawn. The findings should therefore be understood as a synthesis of patterns and associations in the available literature rather than definitive evidence that AI applications improve academic performance directly.

8. Conclusion

This study provides an integrated account of research on artificial intelligence applications and academic performance in higher education by combining bibliometric mapping with a systematic analysis of the included literature. The evidence reviewed indicates that AI applications are increasingly associated with academic support, performance prediction, personalised learning, student engagement, assessment and institutional decision-making. However, direct evidence that AI use improves academic performance remains uneven and should be interpreted in relation to study design, outcome measurement, implementation context and the degree of pedagogical integration. A substantial proportion of the literature evaluates predictive accuracy, technology acceptance, behavioural intention or perceived learning improvement rather than academic outcomes that are sustained and measured objectively.

The study makes four principal contributions. First, it provides an updated mapping of the development of research on AI applications and academic outcomes in higher education during the 2020–2025 period. Second, it distinguishes among different forms of evidence, including technology adoption and perceived usefulness, predictive model performance, correlational relationships and learning or academic outcomes that are measured directly. This distinction prevents the interpretation of technical accuracy or positive perceptions as evidence of educational effectiveness.

Third, the integration of bibliometric and systematic-review methods connects macro-level publication, citation and collaboration patterns with a more detailed examination of research designs, AI applications, educational purposes, outcome measures and theoretical foundations. Fourth, the review identifies the limited use of explicit frameworks that integrate technology, pedagogy, disciplinary content and assessment, despite the presence of technology acceptance, learning, motivational and technical perspectives across the literature.

Several interconnected gaps remain. Methodologically, literature is dominated by predictive, cross-sectional, correlational and self-reported studies, while longitudinal, experimental and intervention-based evaluations remain comparatively limited. Conceptually, AI is treated frequently as a technical, analytical or adoption-related tool without a sufficiently specified explanation of how it should be embedded in teaching and learning practices. Geographically, many studies are restricted to single institutions or national contexts, limiting the transferability of their findings across educational systems, disciplines, languages and levels of digital development.

Taken together, these findings suggest that the educational value of AI should not be assumed from adoption rates, user perceptions or predictive performance alone. Future research should evaluate clearly defined AI-supported interventions using appropriate comparison designs, direct and sustained academic-outcome measures, explicit pedagogical frameworks and diverse institutional contexts. For higher education institutions, responsible implementation requires alignment with learning objectives, faculty preparation, human oversight and appropriate safeguards for academic integrity, privacy, transparency and fairness.

9. Acknowledgements

The authors would like to express their gratitude for the use of Scopus AI to retrieve published studies on the research topic and apply the initial screening criteria, the Napkin AI software for generating the images and figures used, and the Copilot software for writing this paper and for helping to improve the language and grammar. This paper remains an accurate representation of the work and intellectual contributions of the authors. The AI-assisted tools were used only to support language refinement, editing, formatting and visual drafting. The authors conducted all intellectual decisions, screening, coding, analysis, interpretation, framework development and final manuscript approval.

The funding information
The authors state that there is no funding involved.

10. Author contribution

Mohammed Al-Battashi, the corresponding author of this article, supervised the project and wrote the methodology; Hassan Abuhassna did the analysis of the data; Mohammed Azerin drafted the results; all authors worked collaboratively and finalised the last version of the article.

Conflict of interest statements
The authors state that there is no conflict of interest.

Data availability
Data will be made available for a valid reason.

11. References

Abdulayeva, A., Zhanatbekova, N., Andasbayev, Y., Khaimuldanov, Y., & Zhiyembayev, Z. (2025). The role of artificial intelligence (AI) in personalised physics education. *Jurnal Pendidikan IPA Indonesia, 14*(3), 599–615. https://doi.org/10.15294/jpii.v14i3.29752

Alateyyat, S., & Soltan, M. (2024, January). Utilizing artificial intelligence in higher education: A systematic review. In *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)* (pp. 371–374). IEEE. https://doi.org/10.1109/ICETSIS61505.2024.10459629

Altınpulluk, H. (2026). The silent erosion: How AI in education undermines essential skills. In A. Bozkurt (Ed.), *Policy and practice for ethical AI integration in education* (pp. 197–218). IGI Global Scientific Publishing. https://doi.org/10.4018/979-8-3373-7569-4.ch006

Anwar, S., Butt, A. A., & Menekse, M. (2025). Utilizing an NLP-supported mobile reflection application to explore academic engagement, application engagement, and performance in engineering and physics courses. *International Journal of STEM Education, 12*, Article 41. https://doi.org/10.1186/s40594-025-00551-5

Balderrama, L. T., & Araújo Lima, D. (2024). Insights into factors influencing academic success: An application of classification models in higher education. *Artificial Intelligence and Applications, 2*(3), 218–227. https://doi.org/10.47852/bonviewAIA32021465

Bećirović, S., Polz, E., & Tinkel, I. (2025). Exploring students' AI literacy and its effects on their AI output quality, self-efficacy, and academic performance. *Smart Learning Environments, 12*(1), Article 29. https://doi.org/10.1186/s40561-025-00384-3

BinJwair, A. (2025). Predicting STEM students' adoption of generative AI in academic contexts: An application of the UTAUT model. *Frontiers in Education, 10*, Article 1669750. https://doi.org/10.3389/feduc.2025.1669750

Buranasing, A. (2020). Efficiency assessment of undergraduates based on academic record using deep learning methodology. *International Journal of Information and Education Technology, 10*(7), 511–517. https://doi.org/10.18178/ijiet.2020.10.7.1416

Campillo-Ferrer, J. M., López-García, A., & Miralles-Sánchez, P. (2025). Student perceptions of the use of Gen-AI in a higher education program in Spain. *Digital, 5*(3), Article 29. https://doi.org/10.3390/digital5030029

Chavez, H., Chavez-Arias, B., Contreras-Rosas, S., Alvarez-Rodríguez, J. M., & Raymundo, C. (2023). Artificial neural network model to predict student performance using nonpersonal information. *Frontiers in Education, 8*, Article 1106679. https://doi.org/10.3389/feduc.2023.1106679

El Bahri, N., Itahriouan, Z., & Ouazzani Jamil, M. (2025). Emotion-aware education through affective computing and learning analytics: Insights from a Moroccan university case study. *Digital, 5*, Article 45. https://doi.org/10.3390/digital5030045

El Hosayny, O., Razkane, H., & Yeou, M. (2025). Extended TAM-based acceptance of AI for supporting critical thinking in academic research: An examination of perceived benefits and challenges. *Discover Artificial Intelligence, 5*, Article 167. https://doi.org/10.1007/s44163-025-00390-x

Jaboob, M., Al-Ansi, A. M., & Riyadh, H. A. (2025). Artificial intelligence in higher education: A PRISMA 2020 review. *International Journal of Educational Management, 40*(1–2), 113–132. https://doi.org/10.1108/IJEM-09-2024-0560

Kasimovskaya, N., Fomina, E., Morugina, O., Malkina, O., Zhukova, M., & Biryukova, T. (2025). The influence of education using artificial intelligence technologies on the formation of moral behavior and self-efficacy in medical students. *International Journal of Information and Education Technology, 15*(5), 965–976. https://doi.org/10.18178/ijiet.2025.15.5.2303

Krishnan, R., Sowmya, K. R., Wase, D. M., & Kandasamy, L. (2025). Sustainable learning through AI tool adoption among university students in Ethiopia based on an extended UTAUT model. *Asian Journal of University Education, 21*(3), 898–917. https://doi.org/10.24191/ajue.v21i3.60

Kustitskaya, T. A., Esin, R. V., & Noskov, M. V. (2025). Model drift in deployed machine learning models for predicting learning success. *Computers, 14*(9), Article 351. https://doi.org/10.3390/computers14090351

Kwok, A. P. K., Wong, Y. H., Wong, K. C., & Chan, C. H. (2024). AI meeting assistants in English-medium university lectures in Hong Kong, China: A double-edged sword for student perception. *International Journal of Information and Education Technology, 14*(9), 1271–1276. https://doi.org/10.18178/ijiet.2024.14.9.2156

Lakhdar, Y., El-Bendadi, K., & Bakkas, B. (2024). A new hybrid model to predict the performance of trainee teachers based on clustering and classification. *Journal of Computer Science, 20*(9), 1020–1029. https://doi.org/10.3844/jcssp.2024.1020.1029

Ouyang, F., Zheng, L., & Jiao, P. (2022). Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. *Education and Information Technologies, 27*(6), 7893–7925. https://doi.org/10.1007/s10639-022-10925-9

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

Ramírez-Montoya, M. S., Morales-Menendez, R., Tworek, M., Escobar, C. A., Tariq, R., & Tenorio-Sepulveda, G. C. (2024). Complex competencies for leader education: Artificial intelligence analysis in student achievement profiling. *Cogent Education, 11*(1), Article 2378508. https://doi.org/10.1080/2331186X.2024.2378508

Rico-Juan, J. R., Cachero, C., & Macià, H. (2024). Study regarding the influence of a student's personality and an LMS usage profile on learning performance using machine learning techniques. *Applied Intelligence, 54*(8), 6175–6197. https://doi.org/10.1007/s10489-024-05483-1

Tsai, S.-C., Chen, C.-H., Shiao, Y.-T., Ciou, J.-S., & Wu, T.-N. (2020). Precision education with statistical learning and deep learning: A case study in Taiwan. *International Journal of Educational Technology in Higher Education, 17*(1), Article 12. https://doi.org/10.1186/s41239-020-00186-2

Villegas-Ch, W., García-Ortiz, J., Mullo-Ca, K., Sánchez-Viteri, S., & Roman-Cañizares, M. (2021). Implementation of a virtual assistant for the academic management of a university with the use of artificial intelligence. *Future Internet, 13*(4), Article 97. https://doi.org/10.3390/fi13040097

Wilson-Trollip, M. (2025). The effect of artificial intelligence as a peer-to-peer support tool on engagement, grades and pass rates. *Perspectives in Education, 43*(1), 239–266. https://doi.org/10.38140/pie.v43i1.8911

Yan, L., & Yang, Y. (2025). Assessing the learning potential of freshmen in labor education courses using ordinal features and support vector machine. *Frontiers in Education, 10*, Article 1483964. https://doi.org/10.3389/feduc.2025.1483964

Yousef, M., Deeb, S., & Alhashlamon, K. (2025). AI usage among medical students in Palestine: A cross-sectional study and demonstration of AI-assisted research workflows. *BMC Medical Education, 25*(1), Article 693. https://doi.org/10.1186/s12909-025-07272-x

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), Article 39. https://doi.org/10.1186/s41239-019-0171-0