ISRG Journal of Education, Humanities and Literature (ISRGJEHL)

ISRG PUBLISHERS

Abbreviated Key Title: ISRG J Edu Humanit Lit

ISSN: 2584-2544 (Online)

Journal homepage: https://isrgpublishers.com/isrgjehl/

Volume – III Issue – IV (July – August) 2026

Frequency: Bimonthly

Educational Transformation and Governance Turn of Generative Artificial Intelligence in Higher Education: Knowledge Structure, Thematic Evolution, and Research Frontiers

Fu Qiang[1], Bu Bo[2*]

[1, 2] Jilin Engineering Normal University, China.

| **Received:** 19.07.2026 | **Accepted:** 23.07.2026 | **Published:** 25.07.2026

*Corresponding author: Bu Bo

Abstract

Generative artificial intelligence (GenAI) has moved rapidly from a novel conversational technology to an infrastructure-level challenge for higher education. This study maps the development of research on GenAI in higher education and examines whether the field is shifting from an early emphasis on technological adoption and academic disruption toward educational transformation and institutional governance. A bibliometric dataset of 874 Web of Science records published between 2023 and 20 July 2026 was analysed using CiteSpace. Performance indicators, author and institutional collaboration, country participation, document co-citation, keyword co-occurrence, cluster structure, and temporal evolution were examined. The literature expanded from 34 publications in 2023 to 147 in 2024 and 344 in 2025; 349 records had already been indexed by 20 July 2026. The United States, China, England, Australia, and Spain were the most productive contributors, although brokerage centrality was not proportional to output. Co-citation analysis identified a compact intellectual core organised around AI policy, academic integrity, educational opportunities and risks, assessment, and technology acceptance. The keyword network was highly connected (N = 334, E = 1,194; largest component = 95%), with higher education, artificial intelligence, generative AI, academic integrity, and AI literacy as the most frequent terms. Cluster quality was acceptable (Q = 0.5022; weighted mean silhouette = 0.7856), revealing ten themes that were analytically consolidated into adoption and acceptance, teaching and assessment transformation, academic integrity, AI literacy and epistemic judgement, human-AI interaction, measurement development, and governance in diverse institutional settings. The results show a clear but incomplete governance turn: academic integrity and AI literacy have become central, yet technology acceptance models and behavioural intention remain dominant. The field therefore risks treating GenAI primarily as a user-adoption problem while underexamining institutional responsibility, assessment validity, disciplinary variation, equity, and long-term learning. A future research agenda is proposed around observable learning outcomes, task-specific human-AI collaboration, discipline-based AI literacy, policy effectiveness, assessment validity, and substantive educational equity.

Keywords: *generative artificial intelligence; higher education; academic integrity; AI literacy; educational governance; bibliometric analysis; CiteSpace*

DOI: 10.5281/zenodo.21549578

1. Introduction

The public release of ChatGPT at the end of 2022 altered the practical conditions under which higher education produces, communicates, and evaluates knowledge. Earlier forms of artificial intelligence in education were generally embedded in relatively bounded systems, such as intelligent tutoring, predictive analytics, automated scoring, or recommender platforms. Generative artificial intelligence differs because it is directly accessible to teachers and students, responds in natural language, generates extended text and code, and can participate in almost every stage of an academic task. It can help a learner interpret a difficult concept, create examples, formulate questions, draft and revise writing, analyse data, and receive immediate feedback. The same capabilities can also produce fabricated information, conceal weak understanding, displace effort, and blur conventional distinctions between assistance, authorship, and misconduct (Kasneci et al., 2023; Dwivedi et al., 2023; Tlili et al., 2023).

The educational significance of GenAI is therefore not reducible to another cycle of technology adoption. Its capacity to generate plausible academic products destabilises assumptions on which many university practices have depended. A take-home essay once served as a proxy for reading, argumentation, disciplinary understanding, and independent composition. In an environment where a language model can produce a coherent essay in seconds, the submitted text no longer provides straightforward evidence of the student's cognitive contribution. Similar tensions arise in programming, translation, design, lesson planning, feedback, and research writing. The issue is not simply whether AI has been used, but what intellectual work remains attributable to the student, how that work can be made visible, and whether the assessment continues to measure the intended learning outcomes (Cotton et al., 2024; Moorhouse et al., 2023; Perkins, 2023).

Early scholarship responded to this disruption through opportunity-risk inventories, rapid reviews, and commentaries. Studies documented the potential of GenAI to support personalised learning, brainstorming, formative feedback, accessibility, and teacher workload reduction, while warning about hallucinations, bias, privacy, overreliance, and academic dishonesty (Baidoo-Anu & Owusu Ansah, 2023; Farrokhnia et al., 2024; Lo, 2023). This work was necessary because institutions required immediate guidance. Yet the speed of publication also encouraged broad claims based on limited evidence. Much of the early literature discussed what GenAI might do rather than measuring what it actually did in authentic courses, over sustained periods, for different groups of learners.

A second stream of research focused on acceptance and behavioural intention. Technology acceptance models, the unified theory of acceptance and use of technology, trust, habit, performance expectancy, and perceived usefulness quickly became prominent explanatory constructs (Strzelecki, 2024; Venkatesh et al., 2003). Such studies reveal why students and teachers decide to use GenAI and how institutional support or anxiety influences uptake. They are valuable for implementation, but they also frame the central educational question as one of willingness to adopt. Adoption is not equivalent to educational value. A learner may frequently use a chatbot while engaging in shallow verification, outsourcing difficult thinking, or reproducing biased output. Conversely, a cautious user may employ AI selectively in ways that strengthen reflection and disciplinary judgement. The field therefore needs to distinguish use frequency from the quality and epistemic structure of use.

At the same time, academic integrity became a highly visible concern. The initial institutional response often oscillated between prohibition, detection, and cautious permission. Research soon showed that AI-generated text detectors were not a stable solution and that generic prohibitions were difficult to enforce when GenAI functions were embedded in common writing, search, and productivity tools. More durable responses began to emphasise clear disclosure rules, process evidence, oral defence, staged assessment, authentic tasks, and explicit teaching about responsible use (Cotton et al., 2024; Luo, 2024; Moorhouse et al., 2023). This shift reframed academic integrity from an individual compliance issue into a matter of curriculum design, institutional policy, educator capability, and procedural fairness.

The emerging governance perspective is visible in international guidance. UNESCO's guidance for GenAI in education and research promotes a human-centred approach that protects human agency, inclusion, data privacy, and public accountability (UNESCO, 2023). Its later competency frameworks for teachers and students treat AI capability as a combination of human-centred values, ethics, technical understanding, pedagogical application, and responsible creation (UNESCO, 2024a, 2024b). The European Commission similarly links educational use to transparency, fairness, human oversight, privacy, and critical engagement with AI systems (European Commission, 2022). These frameworks indicate that higher education requires more than local rules about cheating. It needs governance arrangements that connect pedagogy, assessment, data, procurement, professional development, and accountability.

Despite the rapid accumulation of publications, it remains unclear how these strands relate within the knowledge structure of the field. Existing reviews frequently map AI in education broadly, combine pre-GenAI and post-ChatGPT research, or concentrate on a narrow application such as writing, language learning, or medical education. Bibliometric studies often report production rankings and keyword clusters but give limited attention to the tension between educational transformation and governance. A large-scale map focused specifically on GenAI in higher education can show whether governance has become a central organising principle or remains peripheral to adoption, perception, and tool-use research.

This study addresses that need by analysing 874 Web of Science records published from 2023 to July 2026. It asks four questions: (1) What are the publication characteristics, influential contributors, and collaboration patterns of GenAI research in higher education? (2) What intellectual foundations and thematic structures constitute the field? (3) How have themes evolved from initial disruption and adoption toward educational transformation and governance? (4) What theoretical, methodological, and institutional directions should guide future research? The article contributes by linking bibliometric structure to an education-centred interpretation. Rather than treating the maps as ends in themselves, it uses them to evaluate the field's underlying problem definition: whether GenAI is being studied mainly as a technology to be accepted, a pedagogical partner to be designed, or a socio-technical system requiring institutional governance.

DOI: 10.5281/zenodo.21549578

2. Conceptual Background

2.1 Generative AI as an educational technology

Generative AI refers to models that produce new content in response to prompts, including text, images, code, audio, and multimodal outputs. In higher education, large language models have become the dominant public interface because they can engage in extended dialogue and adapt explanations to a user's request. Their educational affordances include speed, conversational interaction, scalable feedback, linguistic assistance, simulation of perspectives, and rapid production of examples. These affordances are not fixed educational benefits. They become meaningful only through the relationship among the technology, the task, the learner's prior knowledge, and the teacher's design. The same fluent output can support comparison and critique in one task while enabling unexamined substitution in another.

This distinction is especially important because language models generate statistically plausible sequences rather than verified knowledge. Their outputs may be accurate, partially accurate, fabricated, biased, or internally inconsistent. Fluency can obscure uncertainty. Effective educational use therefore requires learners to inspect sources, test claims, compare outputs with disciplinary evidence, and understand when not to rely on the system. The educational unit of analysis is not the chatbot alone, but the human-AI activity system in which prompts, outputs, judgements, revisions, and disclosures are distributed across people, tools, and institutional rules (Bender et al., 2021; Kasneci et al., 2023).

2.2 Educational transformation

Educational transformation in this study refers to changes in the design and organisation of teaching, learning, assessment, and academic roles. In teaching, GenAI can assist with examples, explanations, lesson resources, formative questions, and differentiated materials. This may reduce routine preparation, but it also requires educators to verify content, preserve disciplinary standards, and design activities in which AI use advances rather than replaces learning. The teacher's expertise shifts from exclusive production of content toward orchestration, judgement, feedback design, and the creation of productive constraints.

In learning, the critical transition is from obtaining answers to regulating a dialogue with an imperfect system. Productive use involves problem formulation, prompt revision, comparison, evidence checking, explanation, and reflective decision-making. These practices can support self-regulation and metacognition when they are explicitly taught. Without such design, the technology may encourage premature closure: learners accept the first plausible output, bypass productive struggle, and confuse linguistic quality with conceptual understanding. The same tool can therefore amplify either agency or dependency.

Assessment is the point at which GenAI most directly exposes weaknesses in inherited practice. Many conventional assignments reward the production of polished text while leaving the process invisible. If the intended construct is argumentation, problem solving, interpretation, or professional judgement, a final product created with opaque AI assistance may not provide valid evidence. Assessment transformation requires tasks that elicit decision processes, disciplinary reasoning, iterative development, oral explanation, and transparent records of AI use. It also requires criteria that evaluate how students question, verify, and improve AI output rather than merely whether AI was present (Moorhouse et al., 2023).

2.3 Governance turn

Governance refers to the allocation of authority, responsibility, rules, and safeguards around GenAI. At the individual level, it includes AI literacy, ethical judgement, disclosure, and self-regulation. At the course level, it includes the alignment of permitted use with learning outcomes, clear instructions, assessment design, and feedback. At the institutional level, it includes policy, staff development, data protection, procurement, accessibility, appeals, and quality assurance. At the system level, it includes legal obligations, professional standards, public values, and the distribution of educational opportunity.

A governance turn occurs when research moves beyond describing benefits and risks to examining how institutions organise responsible practice. It does not imply a movement from permissive use to restriction. Rather, it represents a shift from isolated user behaviour to socio-technical arrangements. A mature governance approach asks who decides which tools may be used, what data may be entered, how students are informed, how decisions can be challenged, how teachers are supported, and how unequal access to premium systems affects assessment. It also asks whether institutional policy is effective in practice, not merely whether a policy document exists.

The analytical logic guiding this review can be summarised as technological affordances leading to educational transformation, which generates new risks and prompts governance responses. The sequence is not linear. Governance shapes how affordances are used, while educational design can prevent or intensify risk. AI literacy, for example, is simultaneously a learning outcome, a condition for responsible use, and a component of institutional governance. This interdependence provides the interpretive frame for the bibliometric results.

3. Methodology

3.1 Research design and data source

The study combined performance analysis and science mapping, following established bibliometric guidance that distinguishes productivity and impact indicators from relational analyses of collaboration, co-citation, and co-word structure (Donthu et al., 2021). Web of Science Core Collection was used because its cited-reference fields support reproducible co-citation analysis and are directly compatible with CiteSpace. The supplied dataset contained 874 records retrieved on 20 July 2026. The publication timespan represented in the final export was 2023-2026, with 2026 treated as an incomplete year.

The search strategy linked GenAI terms with higher education terms using a proximity operator and then required at least one concept associated with academic integrity, ethics, governance, assessment, AI literacy, teacher roles, or human-AI collaboration. The final search string is reproduced in Table 1. This design intentionally excluded broad machine-learning and educational-data-mining terms. It therefore maps the post-ChatGPT literature concerned with educational practice and governance rather than the full history of artificial intelligence in education.

The search produced 874 records. Because the CiteSpace analyses supplied for this study were generated from the complete export, the bibliometric mapping uses all 874 records. Before journal

DOI: 10.5281/zenodo.21549578

submission, the author should retain the original WoS export and document any database-level language or document-type filters in the final methods statement. The present manuscript reports the dataset exactly as used in the supplied maps and avoids claiming additional screening exclusions that cannot be reconstructed from the screenshots alone.

Table 1. Search strategy, dataset, and CiteSpace settings

Component	Specification
Database	Web of Science Core Collection
Retrieval date	20 July 2026
Timespan represented	2023-2026 (2026 incomplete)
Records included	874
Core search logic	TS=(("generative artificial intelligence" OR "generative AI" OR GenAI OR ChatGPT OR "large language model*") NEAR/5 ("higher education" OR "university education" OR "tertiary education")) AND TS=("academic integrity" OR ethic* OR governance OR assessment OR "AI literacy" OR "teacher role*" OR "human-AI collaboration")
CiteSpace version	6.4.R1 (64-bit) Advanced
Time slicing	One-year slices
Selection criteria	g-index (k = 25), LRF = 2.5, L/N = 10, LBY = 5, e = 1.0
Pruning	None
Analyses	Annual production, country, institution, author collaboration, document co-citation, keyword co-occurrence, clustering, and timeline visualisation

3.2 Data standardisation and analysis

The interpretation distinguished literal keyword counts from conceptually equivalent terms. Variants such as generative AI, generative artificial intelligence, GenAI, ChatGPT, large language model, and LLM were treated as related but were not retrospectively collapsed in the supplied network because doing so would change the original node structure. Their simultaneous prominence is itself informative: it indicates rapid terminological formation in a young field. Similarly, academic integrity and academic honesty were interpreted together, while AI literacy, digital literacy, information literacy, and data literacy were kept conceptually distinct.

Four network types were examined. The author collaboration network contained 560 nodes and 2,337 links, with a density of 0.0149 and a largest connected component of 524 nodes (93%). The institution/country-level collaboration map contained 106 nodes and 585 links, with a density of 0.1051 and a largest component of 92 nodes (86%). The document co-citation network contained 493 nodes and 1,931 links, with a density of 0.0159 and a largest component of 467 nodes (94%). The keyword co-occurrence network contained 334 nodes and 1,194 links, with a density of 0.0215 and a largest component of 318 nodes (95%). These large connected components indicate that the field already shares a substantial common knowledge base despite its short history.

Cluster interpretation used the CiteSpace modularity and silhouette statistics. The keyword network produced a modularity Q of 0.5022 and a weighted mean silhouette of 0.7856, with a harmonic mean of 0.6127. A Q value above 0.30 generally indicates a meaningful division of the network, while a silhouette value approaching 1 indicates internal consistency. The solution was therefore considered suitable for interpretation. Automatic labels were retained for transparency, but several were linguistically awkward. The discussion consequently normalises them into educationally meaningful themes without claiming that the software-generated labels are definitive.

Counts and centrality values were transcribed from the supplied CiteSpace tables. Betweenness centrality was interpreted as a brokerage indicator rather than a direct measure of quality. High publication output identifies volume, while high centrality suggests that a country, institution, reference, or keyword connects otherwise separated areas of the network. Co-citation counts were used to identify the intellectual core. Keyword frequency and centrality were used to distinguish established topics from bridging concepts, and the timeline visualisation was used to analyse temporal shifts.

4. Results

4.1 Publication development and leading contributors

The annual distribution demonstrates extraordinary expansion. Only 34 records were indexed in 2023, the first complete publication year after the public release of ChatGPT. Output increased to 147 records in 2024, an increase of 332.4%. It then reached 344 in 2025, representing a further increase of 134.0%. By 20 July 2026, 349 records had already been indexed, slightly exceeding the total for the whole of 2025. Because 2026 is incomplete, its value should not be used as a directly comparable annual growth rate. It nevertheless indicates that the field had not stabilised and was likely to reach a substantially higher year-end total.

The distribution also shows how quickly the literature moved from an emergency response to a major research domain. The 2023 publications formed only 3.9% of the dataset, while 2025 and the partial 2026 period together accounted for almost four-fifths. This concentration in recent years has two implications. Citation indicators are strongly affected by time, so older conceptual articles enjoy a structural advantage. More importantly, the dominant themes of the field are still being formed; clusters that appear peripheral in a 2026 map may become central as empirical and policy research accumulates.

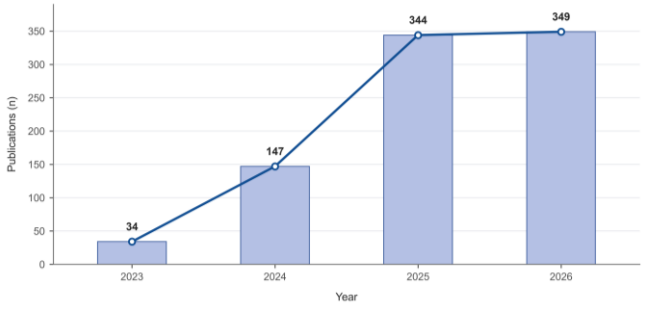

Figure 1. Annual scientific production, 2023-2026. The 2026 count covers records indexed through 20 July 2026.

DOI: 10.5281/zenodo.21549578

Country productivity was led by the United States (130), China (112), England (88), Australia (77), and Spain (70). Germany (32), Saudi Arabia (31), the United Arab Emirates (28), Sweden (28), Türkiye (27), India (26), Malaysia (26), and Mexico (26) formed a second group of active contributors. The distribution is notably international: substantial output came from Europe, North America, East Asia, the Middle East, South Asia, Latin America, and Africa. This breadth reflects the global accessibility of GenAI and the fact that academic integrity and assessment pressures appeared almost simultaneously across higher education systems.

Publication volume did not align perfectly with network brokerage. England showed a centrality of 0.11, higher than the United States (0.01), China (0.06), Australia (0.04), and Spain (0.04). Malaysia had a centrality of 0.12 despite 26 publications, and Brazil reached 0.16 with only 10 publications. Türkiye (0.09), Germany (0.07), Mexico (0.07), Indonesia (0.07), and Egypt (0.08) also occupied bridging positions. These patterns suggest that some medium-output countries connect otherwise separated regional or thematic communities. In a young field, such brokerage may be especially important for transferring policy frameworks, instruments, and empirical designs across contexts.

Institutional output was more dispersed. The University of London was the largest visible contributor with 22 records, followed by the Education University of Hong Kong and the Chinese University of Hong Kong with nine each. Tecnologico de Monterrey, King's College London, the University System of Georgia, and Zayed University each had eight. Deakin University and Monash University each had seven. The appearance of the Egyptian Knowledge Bank with a count of 11 and centrality of 0.20 indicates an affiliation-indexing artefact rather than a conventional research institution. This result underlines the need for affiliation cleaning before producing a final institutional ranking.

Author productivity was low and widely distributed. Jou Min had five records, Zlotnikova Irina had four, and a group of authors including Qu Yao, Hao Yongwei, Zhao Xin, Wang Jue, Chiu Thomas K. F., Cabero-Almenara Julio, Bannister Peter, Chen Xuanning, and Chan Cecilia Ka Yuk each had three. The visible centrality values for the most productive authors were close to zero. The author network nevertheless had a large connected component of 93%, suggesting that many researchers were indirectly linked through co-authorship chains even though no stable small group dominated production. The field is therefore connected but not consolidated around a mature invisible college.

Table 2. Leading countries/regions by publication count

Country/region	Count	Centrality
USA	130	0.01
China	112	0.06
England	88	0.11
Australia	77	0.04
Spain	70	0.04
Germany	32	0.07
Saudi Arabia	31	0.02
United Arab Emirates	28	0.02
Sweden	28	0.05
Türkiye	27	0.09
India	26	0.03
Malaysia	26	0.12

Table 3. Leading institutions visible in the supplied CiteSpace table

Institution	Count	Centrality
University of London	22	0.07
Egyptian Knowledge Bank*	11	0.20
Education University of Hong Kong	9	0.01
Chinese University of Hong Kong	9	0.05
Tecnologico de Monterrey	8	0.00
King's College London	8	0.00
University System of Georgia	8	0.02
Zayed University	8	0.02
University of Illinois System	7	0.00
Deakin University	7	0.02
Monash University	7	0.02

**The Egyptian Knowledge Bank is likely an indexing or affiliation artefact and should be checked against the raw export before submission.*

Table 4. Most productive authors visible in the supplied CiteSpace table

Author or author group	Count
Jou, Min	5
Zlotnikova, Irina	4
Qu, Yao; Hao, Yongwei; Zhao, Xin; Wang, Jue; Chiu, Thomas K. F.; Cabero-Almenara, Julio; Bannister, Peter; Chen, Xuanning; Chan, Cecilia Ka Yuk	3 each

4.2 Collaboration structure and intellectual foundations

Figures 2 and 3 separate national productivity from collaboration structure. Figure 2 shows that output was concentrated in the United States, China, England, Australia, and Spain. Figure 3 presents the CiteSpace collaboration network, which is dense at its centre, where the United States, China, England, Australia, Spain, India, Türkiye, the United Arab Emirates, Malaysia, Canada, Saudi Arabia, and South Africa are closely connected. Large nodes indicate high output, while purple outer rings indicate betweenness centrality. Australia's and Malaysia's visible rings reinforce the distinction between productivity and brokerage. The many small peripheral nodes show broad international participation, but the visual centre remains concentrated among a limited number of national systems.

The collaboration structure is encouraging but also reflects unequal capacity. High-output systems possess strong research universities, established educational technology communities, access to commercial AI tools, and extensive English-language publication networks. Lower-resource settings may experience GenAI as an urgent governance issue without equivalent research infrastructure. The cluster labelled 'underdeveloped academic setting' suggests

DOI: 10.5281/zenodo.21549578

that emerging and under-resourced contexts have become a recognisable theme, yet their role may still be framed through adoption constraints rather than through locally generated theories of educational value and governance.

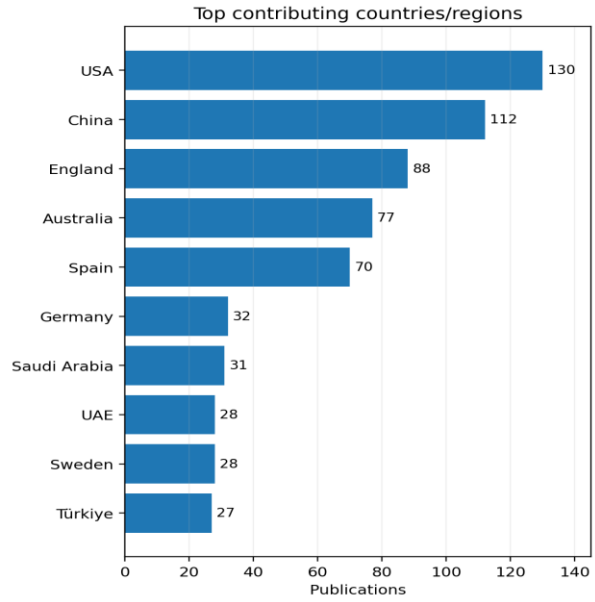

Figure 2. The ten most productive countries and regions by publication count.

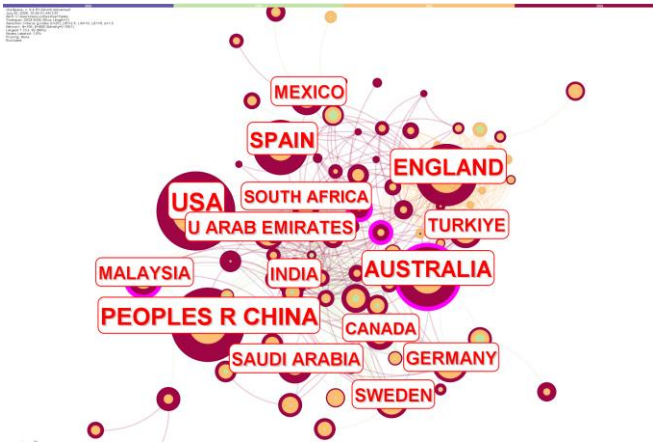

Figure 3. International collaboration network. Node size represents publication output; purple outer rings indicate betweenness centrality.

The document co-citation network contained 493 references and 1,931 links, with 94% of nodes in the largest connected component. This indicates a compact intellectual foundation rather than several isolated literatures. The most frequently co-cited reference was Chan's comprehensive AI policy education framework (179), closely followed by Cotton, Cotton, and Shipway's analysis of academic integrity (177) and Kasneci et al.'s discussion of opportunities and challenges (161). Rudolph, Tan, and Tan's critique of traditional assessment was co-cited 132 times, and Dwivedi et al.'s multidisciplinary opinion paper 121 times. Other highly cited works included Chan and Hu's student-perception study (92), Tlili et al.'s chatbot case study (90), Baidoo-Anu and Owusu Ansah's educational review (89), Lim et al.'s paradoxical perspective on the future of education (81), and Lo's rapid review (71).

The co-citation core is revealing because it combines policy, academic integrity, pedagogical opportunity, assessment disruption, and broad socio-technical reflection. The field did not develop from a single learning theory or educational technology tradition. Instead, it coalesced around a shared problem event: the sudden availability of a powerful generative system. The most influential early publications offered conceptual orientation, institutional frameworks, or rapid synthesis. This explains why policy-oriented and commentary articles are more central than longitudinal learning studies.

Technology acceptance scholarship also forms an important secondary foundation. The co-citation and keyword maps label Venkatesh, Davis, unified theory, behavioural intention, adoption, trust, and the technology acceptance model. These references connect the GenAI literature to established models of information-system adoption. Their presence provides methodological continuity and tested constructs, but it may also narrow the field's explanatory horizon. When acceptance becomes the dominant dependent variable, questions about learning quality, assessment validity, epistemic agency, and institutional responsibility can remain outside the model.

The large nodes for Chan, Cotton, Kasneci, Rudolph, Dwivedi, and Tlili also show how quickly a small group of 2023 articles became canonical. This is partly a first-mover effect. The citation structure should therefore be interpreted as an account of early agenda setting, not a definitive ranking of evidence quality. Newer empirical studies have had less time to accumulate citations, and highly cited conceptual work may be invoked for general context rather than for tested causal claims.

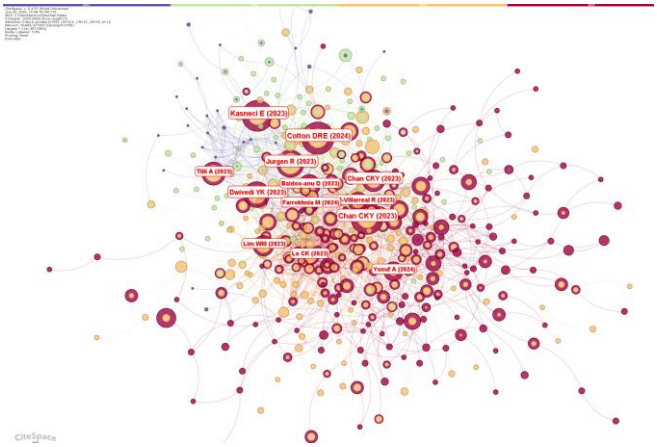

Figure 4. Document co-citation network. Node size represents co-citation frequency; temporal rings indicate the years in which citations occurred.

Table 5. Most frequently co-cited references visible in the supplied CiteSpace table

Rank	Reference (short form)	Co-citations
1	Chan (2023), comprehensive AI policy education framework	179
2	Cotton, Cotton, and Shipway (2024), chatting and cheating	177
3	Kasneci et al. (2023), ChatGPT for good?	161
4	Rudolph, Tan, and Tan (2023), end of traditional assessment?	132

DOI: 10.5281/zenodo.21549578

5	Dwivedi et al. (2023), multidisciplinary perspectives	121
6	Chan and Hu (2023), students' voices on GenAI	92
7	Tlili et al. (2023), guardian angel case study	90
8	Baidoo-Anu and Owusu Ansah (2023), teaching and learning benefits	89
9	Lim et al. (2023), Ragnarok or reformation?	81
10	Lo (2023), rapid review of educational impact	71
11	Farrokhnia et al. (2024), SWOT analysis	67
12	Crompton and Burke (2023), state of the field	64
13	Strzelecki (2024), acceptance and use	61
14	Moorhouse, Yeo, and Wan (2023), assessment guidelines	60
15	Ng et al. (2021), conceptualising AI literacy	59

4.3 Conceptual structure and major themes

The keyword network contained 334 nodes and 1,194 links, with 95% of nodes in one connected component. Higher education was the most frequent keyword (432), followed by artificial intelligence (232), generative AI (230), generative artificial intelligence (193), academic integrity (121), and AI literacy (84). Information technology (56), acceptance (53), user acceptance (45), technology (44), ChatGPT (42), AI (37), critical thinking (37), large language models (32), educational technology (31), technology acceptance model (30), adoption (26), education (26), and students (26) were also prominent.

The prominence of multiple near-synonyms for GenAI is characteristic of an emerging field whose vocabulary has not stabilised. More substantively, the network has two strong organising poles. One concerns educational risk and capability: academic integrity, AI literacy, critical thinking, academic writing, and AI ethics. The other concerns adoption and information systems: acceptance, user acceptance, adoption, technology acceptance, technology acceptance model, unified theory, behavioural intention, trust, and impact. The co-occurrence structure therefore supports the article's central argument that the governance turn is visible but incomplete.

Academic integrity is the most frequent substantive issue after the broad technology and context terms. This confirms that the initial disruption of assessment is not a peripheral subtopic but one of the field's main foundations. AI literacy is also highly prominent and has relatively strong centrality (0.13), indicating that it bridges multiple clusters. AI literacy connects adoption with responsible use, teaching with assessment, and individual capability with institutional policy. It functions as both a research construct and a proposed governance mechanism.

Critical thinking appears 37 times, which is notable because it represents an educational outcome rather than a technology-use construct. However, its position within a network dominated by adoption suggests that critical thinking is often discussed as a hoped-for benefit or risk rather than measured through robust performance evidence. Academic writing (17) is another bridge between learning support and integrity. In this domain, GenAI may assist multilingual students and novice writers, but it can also obscure authorship, reduce practice, and reproduce inaccurate references.

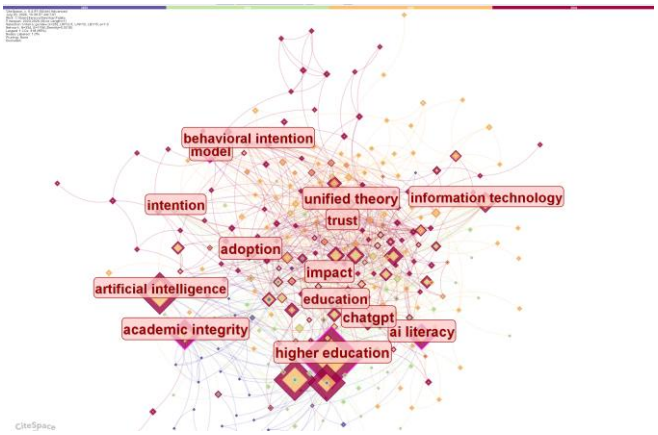

Figure 5. Keyword co-occurrence network. Larger nodes indicate higher frequency; purple rings indicate betweenness centrality.

CiteSpace generated ten clusters. Their automatic labels ranged from clearly interpretable categories such as '#0 ChatGPT adoption' and '#8 addressing academic dishonesty' to less transparent phrases such as '#3 underdeveloped academic setting' and '#6 digital environments evidence'. The overall solution was meaningful (Q = 0.5022) and internally coherent (S = 0.7856). For educational interpretation, the clusters can be consolidated into seven domains: technology adoption and acceptance; qualitative experience and narrative inquiry; ethical and social implications; GenAI in under-resourced academic settings; epistemic beliefs and critical evaluation; measurement and scale development; digital learning evidence; human-AI interaction; and academic integrity. Table 6 retains the automatic labels while providing a normalised interpretation.

The cluster solution also exposes a methodological transition. Adoption, narrative inquiry, and early opportunity-risk discussion represent the field's first phase. Scale development and technology acceptance models mark the conversion of general concern into measurable constructs. Epistemic belief, AI-human interaction, and academic dishonesty point toward more education-specific questions about judgement, agency, and responsibility. The field is therefore moving beyond descriptive reactions, but its institutional governance cluster remains less clearly differentiated than its adoption cluster.

Table 6. CiteSpace clusters and educational interpretation

Cluster	Automatic label	Normalised interpretation	Governance relevance
#0	ChatGPT adoption	Acceptance, intention, trust, and use behaviour	Shows implementation readiness but may over-centre adoption

DOI: 10.5281/zenodo.21549578

#1	Narrative inquiry	Qualitative accounts of student and teacher experience	Reveals meanings, dilemmas, and contextual variation
#2	Technological ethical social	Ethical, social, privacy, bias, and responsibility concerns	Direct foundation for responsible-AI governance
#3	Underdeveloped academic setting	GenAI use in emerging or resource-constrained higher education systems	Highlights infrastructure, capacity, language, and equity
#4	Epistemic belief	Trust, knowledge judgement, critical evaluation, and uncertainty	Connects AI literacy with epistemic agency
#5	Scale development	Measurement of acceptance, literacy, reliance, and perceptions	Supports comparison but requires construct clarity
#6	Digital environments evidence	Learning evidence, digital environments, and educational impact	Calls for observable rather than self-reported outcomes
#7	Mapping AI-human interaction	Human-AI collaboration, co-creation, and division of labour	Central to responsibility and role redesign
#8	Addressing academic dishonesty	Cheating, plagiarism, authorship, disclosure, and detection	Requires assessment and policy reform
#9	Education	General educational integration and broad disciplinary diffusion	Indicates continued expansion beyond early specialist communities

4.4 Thematic evolution and the emerging governance turn

The cluster timeline shows that 2023 was dominated by foundational terms: generative AI, artificial intelligence, higher education, academic integrity, adoption, AI in education, and AI literacy. This was the phase of initial disruption. Research sought to define the technology, identify opportunities and risks, and understand whether students and teachers intended to use it. The large early nodes for generative AI and higher education reflect the rapid formation of the field, while academic integrity became an immediate concern rather than a later consequence.

During 2024, the timeline became more empirical and model-driven. User acceptance, information technology, ChatGPT, students, models, challenges, perceptions, and performance became visible. This period saw the rapid application of TAM and UTAUT-based frameworks, survey instruments, and comparative studies. The shift was useful because it produced systematic evidence on trust, intention, and adoption. Yet the continued prominence of information-technology constructs indicates that GenAI was often treated as a system to be adopted rather than as a force reshaping curriculum and assessment.

By 2025, the visible terms included student engagement, motivation, trust, model, technology acceptance, knowledge, and technology acceptance model. These concepts suggest a move from simple intention to mechanisms and outcomes. At the same time, the persistence of acceptance frameworks shows path dependence. The field was extending existing models rather than consistently developing education-specific theories of human-AI learning. The 2026 nodes suggest continued expansion, but because the year is incomplete and labels are sparse, they should be interpreted cautiously.

The governance turn is therefore best described as emergent rather than complete. Academic integrity, AI literacy, ethics, and AI-human interaction have substantial positions in the network. Policy frameworks and assessment guidelines are among the most co-cited documents. However, governance is not yet the dominant empirical paradigm. Research still concentrates heavily on behavioural intention, acceptance, perception, and self-reported use. Institutional policy effectiveness, appeals and procedural fairness, procurement, privacy practice, unequal access, and long-term assessment validity remain less visible.

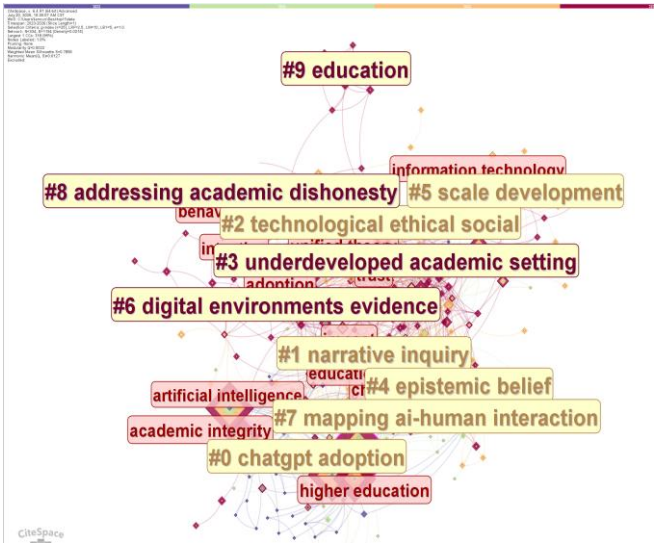

Figure 6. CiteSpace keyword clusters. The labels identify the ten major thematic clusters.

DOI: 10.5281/zenodo.21549578

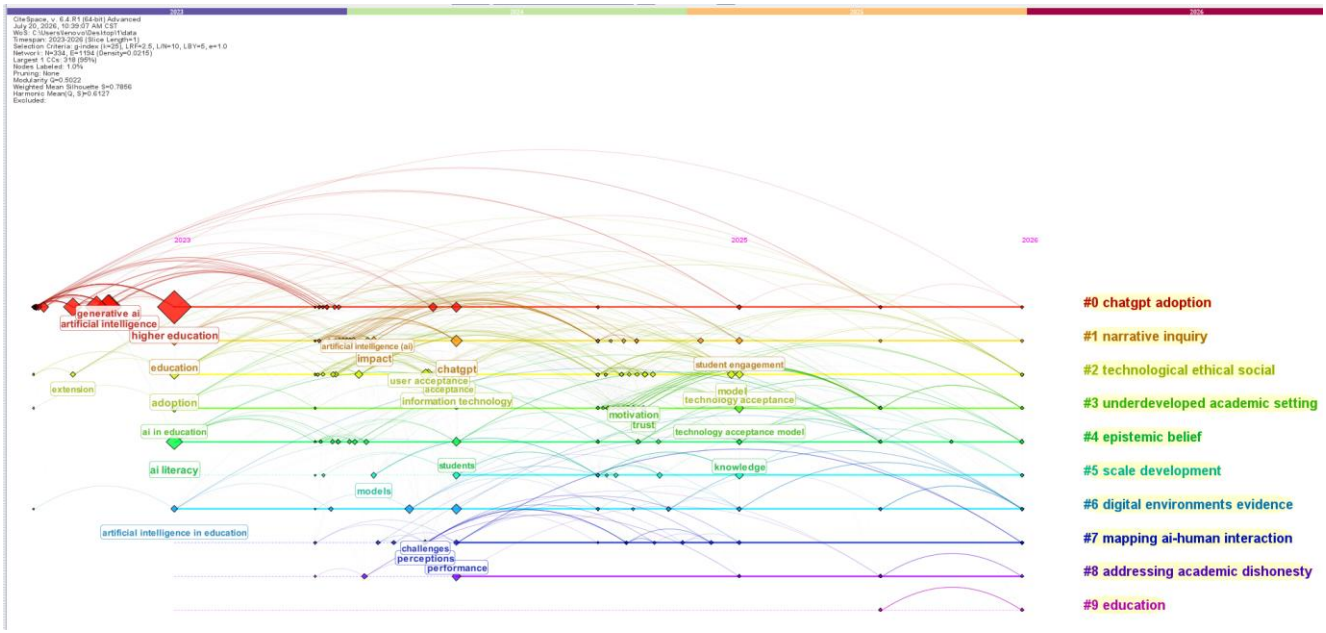

Figure 7. Timeline of thematic clusters, showing their development from 2023 to 2026.

5. Discussion

5.1 Educational Transformation and Human-AI Co-construction

The bibliometric structure confirms that GenAI is not being absorbed into higher education as an ordinary digital tool. The speed of publication and the centrality of assessment, academic integrity, AI literacy, and critical thinking indicate a deeper institutional disturbance. GenAI participates in the production of the very artefacts that universities use to infer learning. It therefore alters both the learning process and the evidentiary basis on which academic judgement is made.

The educational opportunity lies in the technology's capacity to make explanation, feedback, and iteration more available. A student can request alternative explanations, compare arguments, generate examples, or receive immediate suggestions. A teacher can rapidly prototype tasks and resources. These gains are particularly relevant in large classes and multilingual settings. Yet the gains are conditional. Immediate assistance can shorten feedback cycles, but it may also remove the delay and struggle through which learners identify misconceptions and develop durable knowledge. Efficiency is an educational benefit only when it serves the learning objective rather than displacing the activity that the objective requires.

The network's strong adoption orientation reveals a potential category error. Acceptance models explain use, not learning. Performance expectancy may predict intention because students expect faster completion or higher grades, but these outcomes are not equivalent to conceptual growth. Hedonic motivation may increase engagement with the tool without improving disciplinary understanding. Habit may predict use while also signalling dependency. Future studies should therefore separate instrumental success, such as producing an acceptable assignment, from educational success, such as improved transfer, retention, explanation, or judgement.

A more adequate theory of GenAI in higher education must treat tasks as the unit through which affordances become educational. The same student may use AI productively to generate counterarguments but unproductively to write a conclusion without understanding the preceding analysis. The relevant comparison is not simply users versus non-users. It is among patterns of collaboration: student-first then AI feedback, AI-first then student critique, iterative co-construction, AI as tutor, AI as editor, and AI as substitute. Each pattern distributes cognitive work differently and should produce different learning outcomes.

This task-level perspective also connects technology to governance. Institutions cannot write meaningful policy using only broad categories such as 'AI allowed' or 'AI prohibited'. Rules need to specify the purpose, stage, and evidentiary requirements of use. A permitted use in brainstorming may be inappropriate in a task designed to assess independent idea generation. A permitted language-editing function may become problematic when it changes argument content. Educational transformation and governance therefore converge in the design of task-specific boundaries.

Cluster #7 and the growing emphasis on AI-human interaction point toward a new conception of academic work. Human-AI co-construction does not mean that the model and the learner are equivalent partners. The model does not possess educational responsibility, professional accountability, or lived understanding. Co-construction instead describes a workflow in which AI-generated representations become objects for human questioning, selection, verification, and transformation. The quality of the outcome depends on the human ability to direct and evaluate that process.

For teachers, this changes the centre of professional work. Content delivery becomes less exclusive because students can obtain immediate explanations from multiple systems. The teacher's distinctive contribution lies increasingly in designing sequences of activity, identifying misconceptions, creating conditions for dialogue, judging evidence, modelling disciplinary standards, and supporting students when automated responses are misleading. This is not a reduction of the teacher's role. It is an intensification of its interpretive and ethical dimensions.

For students, productive collaboration requires enough prior knowledge to recognise when an answer is implausible. This

DOI: 10.5281/zenodo.21549578

creates a paradox. GenAI is often promoted as support for novices, but novices are least able to evaluate its errors. Without scaffolding, the students who most need assistance may also be most vulnerable to confident misinformation. AI literacy therefore cannot be taught as generic prompt technique. It must be integrated with disciplinary knowledge, source evaluation, uncertainty, and the ability to justify why an output was accepted, revised, or rejected.

The cluster labelled epistemic belief is important in this respect. GenAI use is an epistemic practice: learners decide what counts as knowledge, whose authority to trust, how to resolve conflict, and when evidence is sufficient. Research should examine whether AI interaction encourages sophisticated epistemic judgement or a transactional view in which knowledge is treated as an immediately available answer. Measures of trust should distinguish calibrated trust from general confidence. The desired outcome is not low trust or high trust, but trust that changes appropriately with task risk, evidence quality, and model limitations.

Human-AI co-construction also has a social dimension. Premium models, institutional licences, language support, and digital infrastructure influence who can engage in more productive forms of collaboration. Students with stronger prior knowledge, better prompts, paid access, and experienced teachers may derive greater benefit, while others use less capable tools or lack guidance. Without equity-sensitive design, GenAI may widen differences while appearing universally accessible.

5.2 Assessment Reform and Academic Integrity Governance

Assessment is the clearest point at which the field's educational and governance concerns meet. The co-citation core places Cotton et al., Rudolph et al., Moorhouse et al., and Perkins among the most influential sources, and academic integrity is the fifth most frequent keyword. The initial response to GenAI often treated the submitted product as suspicious and sought technical detection. This approach is unstable because model outputs change, paraphrasing is easy, and detectors can produce false positives. It also places students in a procedurally weak position when accusations rely on opaque probability scores.

The deeper problem is assessment validity. If an assignment can be completed by a model without the intended cognitive process, the institution should ask whether the task still provides valid evidence of learning. Returning entirely to invigilated examinations may protect authorship, but it can also narrow the range of outcomes assessed and disadvantage students for whom authentic professional tasks involve tools, collaboration, and extended work. The appropriate response is not a single AI-proof format. It is a portfolio of assessment modes that balances authenticity, process visibility, independent performance, and responsible tool use.

Several design principles follow. Complex tasks should be staged so that students submit problem definitions, source choices, intermediate analyses, drafts, and reflective revisions. Oral defence can test ownership and reasoning. In-class components can establish a baseline of independent capability. Authentic projects can require local data, personal decisions, or context-specific evidence that generic models cannot supply reliably. Students can be required to annotate where AI was used, preserve relevant prompts and outputs, and explain how they verified and changed the generated material.

Assessment criteria also need reform. A rubric for AI-supported work should not reward the superficial polish of the final text alone. It should evaluate problem framing, source quality, verification, disciplinary reasoning, the identification of model limitations, ethical judgement, and the student's contribution to revision. In some tasks, the ability to critique a weak AI response may be more educationally valuable than producing a response without AI. Such design turns GenAI from an invisible substitute into an explicit object of learning.

The field should also distinguish assessment literacy from detection literacy. Teachers need to understand what current systems can and cannot do, but the goal is not to become forensic investigators. Their primary competence is to design tasks that elicit valid evidence and to communicate rules consistently. Institutions should support this work through workload allocation, exemplars, moderation, and professional development. Otherwise, governance policy will place responsibility on individual instructors without giving them the resources to implement it.

The cluster on academic dishonesty should not be interpreted as proof that GenAI has transformed integrity into an entirely new problem. Contract cheating, unauthorised collaboration, plagiarism, and fabrication predate large language models. What GenAI changes is scale, accessibility, ambiguity, and the ease with which assistance can become substitution. It also challenges policies built around textual similarity because generated text may be original in wording while still misrepresenting authorship or learning.

A multi-level approach is needed. At course level, teachers should specify whether GenAI is prohibited, permitted with conditions, or required; identify the stages at which it may be used; state disclosure expectations; and explain how compliance will be evidenced. At institutional level, policies should use common categories so that students are not confronted with contradictory meanings across courses. Institutions need fair investigation procedures, guidance on the limits of detector evidence, data-protection rules, and support for staff who redesign assessment.

At system level, governance must address responsibility and market power. Commercial AI providers determine model updates, data retention, pricing, and functionality. Universities may rely on tools whose operation is not fully transparent and whose paid versions produce different results from free versions. Procurement decisions are therefore educational decisions. Contracts should address data use, accessibility, auditability, service continuity, and the protection of student intellectual property. Governance that focuses only on student misconduct misses these institutional responsibilities.

The most defensible direction is a transition from prohibition-centred policy to transparent, differentiated, and educative policy. Prohibition remains appropriate where independent performance is itself the outcome or where confidentiality makes external tools unsafe. Permission is appropriate where AI use reflects professional practice or supports learning. Between these positions lies a wide range of conditional uses. Governance quality depends on making these distinctions explicit and aligning them with educational purpose.

5.3 Critical AI Literacy and Institutional Capacity

AI literacy is one of the strongest bridging concepts in the network, with 84 occurrences and centrality of 0.13. Its prominence reflects

DOI: 10.5281/zenodo.21549578

recognition that responsible GenAI use cannot be secured by rules alone. Students and teachers need capabilities to understand, operate, evaluate, and govern AI. However, the construct risks becoming too broad. A scale that combines prompt skill, technical knowledge, ethical attitudes, confidence, and frequency of use may produce a convenient score without clarifying which capability predicts which outcome.

Critical AI literacy should include at least five dimensions: understanding how generative systems produce output; evaluating accuracy, evidence, and uncertainty; recognising bias, privacy, and power; managing one's own reliance; and taking responsibility for disclosure and consequences. These dimensions are related but not interchangeable. A student may write effective prompts while failing to verify citations. Another may understand bias conceptually but still over-rely on AI under time pressure. Measurement development should preserve these distinctions.

Disciplinary AI literacy is equally important. In engineering, students must connect AI-generated calculations or designs to safety, standards, and professional responsibility. In medicine, the consequences of hallucination are tied to patient safety and clinical judgement. In teacher education, students must learn not only to use AI for their own study but also to design future classroom use, protect children's data, and evaluate pedagogical appropriateness. In humanities, questions of voice, interpretation, and authorship are central. A single generic literacy framework cannot fully represent these different epistemic and ethical demands.

The cluster labelled scale development indicates that the field is beginning to operationalise AI literacy, acceptance, trust, and reliance. This is a positive move, but new measures should be validated against behaviour and performance rather than only correlated with self-report constructs. Scenario-based judgement tests, trace data from AI interactions, think-aloud protocols, and discipline-specific tasks can reveal whether reported literacy translates into responsible practice.

At the institutional level, the bibliometric evidence supports an integrated framework with five connected components: policy clarity, assessment redesign, capability development, transparency and accountability, and equity and data protection. Policy clarity establishes the categories and boundaries of use. Assessment redesign changes the incentives and evidence structure surrounding student work. Capability development equips students and staff to act within the policy. Transparency and accountability make AI participation visible and contestable. Equity and data protection define the conditions under which institutional adoption remains legitimate.

These components should not be implemented as separate initiatives. A disclosure rule without assessment redesign may produce formal compliance while leaving the task vulnerable to substitution. Assessment redesign without staff development may create excessive workload and inconsistent practice. AI literacy programmes without clear policy may teach capability while leaving students uncertain about permission. Procurement without equity analysis may create a two-tier learning environment in which some students access stronger models. Governance maturity depends on alignment.

Institutions should therefore develop governance through iterative evidence. Policies should be reviewed using data on student understanding, staff implementation, integrity cases, appeals, learning outcomes, and differential access. Discipline-level adaptation is essential, but local variation should sit within a common institutional language. Students should be involved in policy design because they experience conflicting rules and can reveal how guidance is interpreted in practice. Libraries, learning developers, disability services, IT security, legal teams, and academic integrity offices should participate alongside teaching staff.

Governance should also preserve educational experimentation. Overly rigid central rules may prevent teachers from testing productive uses, while completely devolved rules produce inconsistency. A tiered model can combine institution-wide principles with programme and course specifications. High-risk uses involving sensitive data, automated decisions, or professional safety require stronger oversight. Lower-risk uses such as brainstorming or language feedback can be managed through disclosure and task design.

5.4 Structural Gaps and Future Research Agenda

The maps reveal several structural weaknesses. Geographically, the field is broad but led by a relatively small group of countries. Publication participation should not be mistaken for equal agenda-setting power. English-language journals, well-resourced universities, and commercial-tool access influence which problems become visible. Research in lower-resource settings may focus on barriers and acceptance while underrepresenting local pedagogies, multilingual use, public infrastructure, and different regulatory environments.

Methodologically, technology acceptance and cross-sectional perception studies are overrepresented. These designs are efficient and publishable, but they offer limited evidence about learning. Self-reported intention does not establish actual use, and self-reported benefit does not establish improved performance. Short interventions may show immediate gains while missing effects on retention, independence, or transfer. The field needs longitudinal, quasi-experimental, multi-institutional, and mixed-method research grounded in authentic courses.

Outcome selection is also unbalanced. Perceived usefulness, satisfaction, intention, and attitudes are common, while disciplinary reasoning, durable knowledge, creativity quality, judgement under uncertainty, and the ability to detect model error are less frequently measured. Research should examine both benefits and displacement effects. A tool may improve the immediate quality of a submitted product while reducing opportunities to practise the skill the product is intended to demonstrate.

Stakeholder coverage remains incomplete. Students dominate the literature, while teachers, programme leaders, assessors, learning designers, academic integrity investigators, librarians, disability professionals, and institutional leaders are less visible. Governance cannot be understood from student surveys alone. It requires analysis of how policy is designed, communicated, enacted, contested, and revised across organisational roles.

Against these structural limitations, the next phase of research should move from asking whether GenAI is accepted to identifying the conditions under which particular forms of use improve learning. Studies should specify the task, model, prompt process, student knowledge, teacher guidance, and assessment context. They should compare collaboration patterns rather than treating all AI use as one exposure. Where possible, outcome measures should

DOI: 10.5281/zenodo.21549578

include independent post-tests, delayed retention, transfer, error detection, explanation quality, and process traces.

Policy research should examine effectiveness rather than document presence. Comparative studies can test whether prohibition, permission, or differentiated guidance produces clearer understanding, fairer enforcement, and better learning. Institutional policies should be analysed alongside course-level implementation because formal rules may diverge from classroom practice. Integrity cases and appeals can provide important evidence, provided privacy is protected.

Equity research should extend beyond access. Free and paid models differ in capability, reliability, and limits. Students vary in language, disability, prior knowledge, time, and access to expert guidance. Research should evaluate whether GenAI narrows or widens gaps in performance and agency, and whether institutional licences or support programmes change these effects. Global South contexts should be included as sites of theory development rather than only as settings for replicating acceptance models.

Finally, the field needs stronger conceptual integration. AI literacy, epistemic cognition, self-regulated learning, assessment validity, and socio-technical governance should not develop as separate literatures. A mature account of GenAI in higher education must explain how individual capability interacts with task design, disciplinary norms, institutional policy, and commercial infrastructure. Table 7 translates these needs into a focused research agenda.

Table 7. Future research agenda for GenAI in higher education

Research gap	Priority question	Recommended design	Expected contribution
Self-reported perceptions dominate	When and for whom does GenAI improve durable learning, transfer, and independent performance?	Longitudinal and quasi-experimental studies with delayed post-tests	Separates product improvement from genuine learning
AI use treated as a single category	How do student-first, AI-first, critique, tutoring, editing, and substitution workflows differ?	Task-level experiments and process-trace analysis	Builds a theory of human-AI division of cognitive labour
Generic AI literacy constructs	What capabilities are required in engineering, medicine, teacher education, business, and humanities?	Discipline-specific instrument development with performance validation	Connects literacy to professional epistemology and ethics
Policy presence studied more than policy effect	Which policy models improve clarity, fairness, learning, and consistent implementation?	Cross-institutional comparative and implementation research	Moves governance from principle to evidence
Integrity framed as detection	Which assessment designs remain valid when GenAI is permitted?	Assessment-design trials, oral defence, staged submissions, authentic tasks	Reorients integrity toward validity and process evidence
Access treated as binary	Do model tier, language, disability, prior knowledge, and institutional support create unequal benefits?	Equity-focused mixed methods and subgroup analysis	Identifies substantive rather than nominal access
Student perspective dominates	How do teachers, leaders, integrity staff, librarians, and support services enact governance?	Multi-stakeholder case studies and organisational ethnography	Explains policy-practice gaps and institutional capacity
Short-term studies dominate	How does sustained GenAI use affect autonomy, reliance, critical thinking, and professional identity?	Semester- and programme-level longitudinal studies	Tests cumulative benefits and displacement effects

6. Conclusion

This bibliometric review mapped 874 Web of Science records on GenAI in higher education published from 2023 to 20 July 2026. The field expanded at exceptional speed, with 2026 output already exceeding the full 2025 total by the retrieval date. Research participation is global, although the United States, China, England, Australia, and Spain lead publication output. Collaboration networks are highly connected, but author productivity remains dispersed and institutional data require careful cleaning.

The intellectual core was established by early work on AI policy, academic integrity, assessment disruption, educational opportunities and risks, and technology acceptance. The conceptual network confirms that academic integrity and AI literacy are central, while critical thinking, ethics, academic writing, and human-AI interaction provide bridges toward education-specific questions. At the same time, acceptance, behavioural intention, trust, TAM, and UTAUT remain dominant. The central finding is therefore a governance turn that is real but incomplete.

The future of GenAI in higher education will not be determined only by model capability or user willingness. It will depend on whether institutions can redesign tasks and assessments, cultivate critical and disciplinary AI literacy, make human and AI contributions transparent, protect data and equity, and evaluate policy through evidence. The central research question should shift from whether students and teachers will adopt GenAI to how human-AI activity can be organised so that technology expands learning without weakening intellectual agency, assessment validity, or educational responsibility.

DOI: 10.5281/zenodo.21549578

7. Limitations and Future Research

The study has several limitations. It uses a single Web of Science export and therefore does not include publications indexed only in Scopus, ERIC, Dimensions, regional databases, books, or conference proceedings. English-language indexing practices may underrepresent research in other languages. The dataset begins in 2023 and covers only a short period of rapid change. Citation counts favour early publications, while the 2026 publication total is incomplete.

The analysis was conducted from CiteSpace outputs supplied by the author rather than from the raw WoS file within this manuscript-generation workflow. Consequently, the paper reports the visible network settings and tables but does not independently reproduce preprocessing decisions. The institutional list contains at least one likely indexing artefact, demonstrating the importance of affiliation normalisation. Before submission, the raw export should be archived, author and institution names should be harmonised, and document-type and language filters should be stated exactly.

Bibliometric maps reveal patterns of association, not causal relationships or direct measures of research quality. Automatic cluster labels are algorithmic summaries and require substantive interpretation. The proposed governance framework and future research agenda are therefore analytical syntheses derived from the mapped literature, not tested institutional interventions. Future updates should combine WoS and Scopus, include multilingual research, examine citation-normalised influence, and conduct structured content coding of representative studies within each cluster.

Declarations

Funding

No specific funding information was provided for this manuscript. Replace this statement with the relevant funding acknowledgement before submission.

Data availability

The bibliographic data were retrieved from the Web of Science Core Collection. The cleaned export and CiteSpace project files may be made available by the corresponding author subject to database licensing conditions.

Ethics approval

This study analysed bibliographic metadata from published literature and did not involve human participants, personal data collection, or animal research. Ethical approval was therefore not required.

Competing interests

The author declares no competing interests.

References

1. Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-T
2. Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 11(4), 959-975. https://doi.org/10.1016/j.joi.2017.08.007
3. Baidoo-Anu, D., & Owusu Ansah, L. (2023). Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Journal of AI, 7(1), 52-62. https://doi.org/10.61969/jai.1337500
4. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610-623). https://doi.org/10.1145/3442188.3445922
5. Bhullar, P. S., Joshi, M., & Chugh, R. (2024). ChatGPT in higher education - A synthesis of the literature and a future research agenda. Education and Information Technologies. https://doi.org/10.1007/s10639-024-12723-x
6. Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. International Journal of Educational Technology in Higher Education, 20, 38. https://doi.org/10.1186/s41239-023-00408-3
7. Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: Perceptions, benefits, and challenges in higher education. International Journal of Educational Technology in Higher Education, 20, 43. https://doi.org/10.1186/s41239-023-00411-8
8. Chan, C. K. Y., & Lee, K. K. W. (2023). The AI generation gap: Are Gen Z students more interested in adopting generative AI such as ChatGPT in teaching and learning than their Gen X and Millennial generation teachers? Smart Learning Environments, 10, 60. https://doi.org/10.1186/s40561-023-00269-3
9. Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. Journal of Science Education and Technology, 32, 444-452. https://doi.org/10.1007/s10956-023-10039-y
10. Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148
11. Crawford, J., Cowling, M., & Allen, K. A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence. Journal of University Teaching & Learning Practice, 20(3). https://doi.org/10.53761/1.20.3.02
12. Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. International Journal of Educational Technology in Higher Education, 20, 22. https://doi.org/10.1186/s41239-023-00392-8
13. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319-340. https://doi.org/10.2307/249008
14. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285-296. https://doi.org/10.1016/j.jbusres.2021.04.070
15. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A.,

DOI: 10.5281/zenodo.21549578

Raghavan, V., Ahuja, M., Albanna, H., et al. (2023). Opinion paper: So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

16. Eke, D. O. (2023). ChatGPT and the rise of generative AI: Threat to academic integrity? Journal of Responsible Technology, 13, 100060. https://doi.org/10.1016/j.jrt.2023.100060
17. European Commission. (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators. Publications Office of the European Union. https://doi.org/10.2766/153756
18. Farrokhnia, M., Banihashem, S. K., Noroozi, O., & Wals, A. (2024). A SWOT analysis of ChatGPT: Implications for educational practice and research. Innovations in Education and Teaching International, 61(3), 460-474. https://doi.org/10.1080/14703297.2023.2195846
19. Kasneci, E., Sessler, K., Kuechemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Guennemann, S., Huellermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274
20. Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(2), 537-550. https://doi.org/10.1177/00336882231162868
21. Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarok or reformation? A paradoxical perspective from management educators. The International Journal of Management Education, 21(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790
22. Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), 410. https://doi.org/10.3390/educsci13040410
23. Luo, J. (2024). A critical review of GenAI policies in higher education assessment: A call to reconsider the 'originality' of students' work. Assessment & Evaluation in Higher Education. https://doi.org/10.1080/02602938.2024.2309963
24. Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. Computers and Education Open, 5, 100151. https://doi.org/10.1016/j.caeo.2023.100151
25. Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041
26. 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., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71
27. Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. Journal of University Teaching & Learning Practice, 20(2), Article 7. https://doi.org/10.53761/1.20.02.07
28. Rudolph, J., Tan, S., & Tan, S. (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? Journal of Applied Learning & Teaching, 6(1). https://doi.org/10.37074/jalt.2023.6.1.9
29. Strzelecki, A. (2024). To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. Interactive Learning Environments, 32. https://doi.org/10.1080/10494820.2023.2209881
30. Sullivan, M., Kelly, A., & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. Journal of Applied Learning & Teaching, 6(1), 1-10. https://doi.org/10.37074/jalt.2023.6.1.17
31. Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 10, 15. https://doi.org/10.1186/s40561-023-00237-x
32. UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.
33. UNESCO. (2024a). AI competency framework for students. UNESCO.
34. UNESCO. (2024b). AI competency framework for teachers. UNESCO.
35. van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538. https://doi.org/10.1007/s11192-009-0146-3
36. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425-478. https://doi.org/10.2307/30036540
37. Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltynek, T., Guerrero-Dib, J., Popoola, O., Sigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19, 26. https://doi.org/10.1007/s40979-023-00146-z
38. Zawacki-Richter, O., Marin, 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, 39. https://doi.org/10.1186/s41239-019-0171-0

DOI: 10.5281/zenodo.21549578