Black Sea Journal of Engineering and Science
doi: 10.34248/bsengineering.1934728

Open Access Journal
e-ISSN: 2619 8991

Research Article
Volume 9 - Issue 5: 2692-2717 / September 2026

TRENDS AND BIBLIOMETRIC MAPPING OF THE KNOWLEDGE STRUCTURE OF GENERATIVE ARTIFICIAL INTELLIGENCE RESEARCH IN HIGHER EDUCATION

Şener BALAT[1]*

[1]*Bingöl University, Vocational School of Information Technologies, 12000, Bingöl, Türkiye*

Abstract: Generative artificial intelligence (GenAI) has reshaped teaching, learning and assessment in higher education faster than the literature describing it has been consolidated. Existing bibliometric reviews of this area are typically built on a single database, organised around ChatGPT alone, or extended to education at all levels and they rarely report the analytical parameters required for replication. This study addresses that gap by mapping the scientific production structure, intellectual and conceptual organisation and thematic evolution of GenAI research in higher education on a dual-database (Web of Science Core Collection and Scopus), PRISMA 2020-screened and fully parameter-transparent corpus of 2,992 articles and reviews published between 2022 and 2025, drawn from 728 sources and 9,948 authors. Descriptive, keyword co-occurrence, correspondence-analysis-based conceptual structure, strategic (thematic) map and thematic evolution analyses were triangulated across Bibliometrix/Biblioshiny and VOSviewer. Output grew from 5 documents in 2022 to 2,001 in 2025 a 400-fold expansion, with 66.9% of the entire corpus published in the final year alone at a mean of 18.46 citations per document. Impact is steeply concentrated: the ten most-cited documents (0.33% of the corpus) account for 8.75% of all citations. Production is led by China (n = 890), the United States (n = 617) and Australia (n = 401), yet international co-authorship stands at only 10.11%, indicating a field that is collaborative but overwhelmingly domestic. Türkiye (n = 96) ranks in the same output band as the United Arab Emirates, Peru and Hong Kong, but below regional peers such as Saudi Arabia, Indonesia and Jordan. Two independent clustering procedures each return a four-cluster architecture, but not the same four: the co-occurrence network built on indexed vocabulary separates the field by disciplinary setting instructional technology, health-professions education, technology acceptance and a general artificial intelligence core whereas correspondence analysis on author keywords separates it by conceptual concern into technological, educational-application, ethical-assessment and pedagogical dimensions. Generative AI is the single motor theme and the field shifts clearly from technological exploration towards pedagogical integration and academic integrity. Beyond describing the field, the study contributes a fully reproducible bibliometric protocol and a critical reading of the field's ChatGPT-centrism, its collaboration deficit and its early signs of thematic saturation.

Keywords: Generative artificial intelligence, Higher education, Bibliometric analysis, Research trends, Knowledge mapping

***Corresponding author:** Bingöl University, Vocational School of Information Technologies, 12000, Bingöl, Türkiye
Email: sbalat@bingol.edu.tr (Ş. BALAT)
Şener BALAT https://orcid.org/0000-0002-9683-1778

Received: April 20, 2026
Accepted: August 24, 2026
Published: September 15, 2026

Cite as: Balat, Ş. (2026). Trends and bibliometric mapping of the knowledge structure of generative artificial intelligence research in higher education. Black Sea Journal of Engineering and Science, 9(5): 2692-2717.

1. Introduction

The rapid advancement of artificial intelligence technologies has led to a significant paradigm shift in the global technological ecosystem, profoundly influencing how societies produce, process and disseminate knowledge. At the center of this transformation is generative artificial intelligence, which refers to systems capable of producing content such as text, images, code and multimedia through patterns learned from large datasets (Ng and Ho, 2025). While traditional artificial intelligence applications have largely been based on classification or prediction, generative artificial intelligence systems particularly through large language models (LLMs) offer advanced capabilities in understanding and generating natural language (Dúo-Terrón, 2024). The emergence of models such as ChatGPT demonstrates the potential of artificial intelligence to assume a more active role in knowledge production and learning processes (Francis et al., 2025). The development of large language models has attracted considerable attention across multiple disciplines due to their ability to simulate human-like reasoning processes, generate personalized responses and support complex cognitive tasks. These advancements present both significant opportunities and various challenges for modern digital infrastructures (Bittle and El-Gayar, 2025). The integration of generative artificial intelligence tools into professional and academic processes signals a substantial shift in how information is accessed and utilized. The transition from keyword-based information retrieval to generative synthesis-based approaches is reshaping individuals' interactions with knowledge. This

This work is licensed (CC BY-NC 4.0) under Creative Commons Attribution 4.0 International License

transformation necessitates a reconsideration of digital literacy and cognitive engagement concepts (Dai et al., 2026).

The impact of generative artificial intelligence technologies has generated extensive debate, particularly within the context of higher education. Universities are traditionally regarded as central institutions for knowledge production and critical thinking. However, the widespread adoption of these technologies has raised the need for restructuring teaching, learning and assessment processes. Research indicates that generative artificial intelligence tools have the potential to provide personalized learning experiences and contribute to learning environments that adapt to students' individual needs (Bittle and El-Gayar, 2025; Francis et al., 2025).

The influence of these technologies extends beyond instructional processes. Generative artificial intelligence systems can affect students' learning motivation and perceptions of self-efficacy (Guo et al., 2026). They can also support academic writing processes by facilitating idea generation and text production. Nevertheless, the use of generative artificial intelligence in higher education is often characterized as a double-edged phenomenon. While it supports learning processes, it also poses risks related to academic integrity and originality (Bittle and El-Gayar, 2025; Francis et al., 2025).

Assessment and feedback processes are among the areas most affected by these technologies. The ability of large language models to generate academically sophisticated texts has raised concerns that students may complete assignments with limited engagement in the learning process. Consequently, some studies identify this issue as a potential risk to academic integrity in higher education (Lodge et al., 2023). In response to these developments, many universities are developing new policies and pedagogical approaches aimed at both leveraging opportunities and maintaining academic standards (Plata et al., 2023; Chiu, 2024). In this context, the development of pedagogical approaches that support the sustainable and ethical use of artificial intelligence technologies has become increasingly important (Chiu, 2024).

Research on generative artificial intelligence in higher education has shown a notable increase, particularly after 2022 (Ng and Ho, 2025; Samala et al., 2025; Dai et al., 2026). This body of literature has evolved around various research themes. Early studies primarily focused on technical infrastructures and applications within the context of educational technologies (Ng and Ho, 2025). Subsequent research has examined the adoption of these technologies through theoretical frameworks such as technology acceptance models (Guo et al., 2026).

Another prominent theme in the literature concerns academic integrity and ethical issues. The impact of generative artificial intelligence on assessment processes has necessitated the development of new regulations within universities (Plata et al., 2023; Francis et al., 2025). Recent studies have increasingly focused on the effects of these technologies on students' cognitive engagement and higher-order thinking skills (Ng and Ho, 2025). Additionally, pedagogical innovation and the reconsideration of teaching strategies have emerged as important areas of research (Chiu, 2024; Francis et al., 2025).

Despite the rapid increase in studies on generative artificial intelligence and higher education, the literature still exhibits a fragmented structure (Dai et al., 2026). A significant portion of existing research consists of qualitative studies focusing on specific contexts (Prayuda et al., 2026), which makes it difficult to comprehensively understand the overall developmental dynamics of the field. Moreover, there is a limited number of systematic investigations examining publication networks, influential researchers and foundational studies. The conceptual relationships among research themes and their evolution over time have also not been sufficiently explored (Dai et al., 2026). Therefore, there is a clear need for studies that comprehensively analyze the intellectual and conceptual structure of the field.

A closer reading of the bibliometric work already available shows that the gap is one of scope and transparency rather than of quantity. Bozkurt (2023) mapped generative AI, conversational agents and chatbots across education as a whole rather than higher education specifically; Nugroho (2024) organised its corpus around ChatGPT alone; Kavitha et al. (2024) and Flores-Velásquez et al. (2024) addressed artificial intelligence in general rather than its generative variant, the latter narrowing further to technology acceptance; Dúo-Terrón (2024), Ng and Ho (2025) and Guo et al. (2026) examined generative artificial intelligence in education across all levels; and Prayuda et al. (2026) traced two decades of artificial intelligence in education, a window in which generative artificial intelligence occupies only the final three years. Dai et al. (2026) is the closest antecedent, being both recent and specific to higher education, but its emphasis falls on emerging trends and global power dynamics. What is missing across this body of work is a higher-education-specific map that is simultaneously (i) built on the two major citation databases in combination rather than on one, (ii) reported at a level of parameter detail that permits replication and (iii) cross-validated by more than one clustering algorithm.

This omission is consequential, because bibliometric cluster solutions are sensitive to precisely the choices that typically go unreported. Whether keywords are drawn from author keywords or from Keywords Plus, whether the minimum occurrence threshold is set at three or at ten and whether clustering proceeds by modularity optimisation or by factorial projection can each alter the number and the composition of the clusters obtained (van Eck and Waltman, 2010; Cobo et al., 2011; Zupic and Čater, 2015; Aria and Cuccurullo, 2017). A map of generative artificial intelligence research in higher education that does not disclose these settings

cannot be audited, replicated or extended by anyone else. In a field growing at the rate documented in this study, an unverifiable map is of limited value to the researchers and policymakers who are its intended audience.

Why, then, is this study necessary? Three reasons can be stated concretely. First, research on generative artificial intelligence in higher education has expanded so rapidly since late 2022 that the field has outgrown the descriptive maps produced during its first eighteen months and no consolidated account exists of where it now stands. Second, the accumulated literature has not previously been examined simultaneously across the two databases with the widest complementary coverage, so the available picture is partial by construction. Third, the structure of the field has not been established by a procedure that another researcher can reproduce. This study responds to each of these points directly. It (i) builds a dual-database, PRISMA 2020-screened corpus; (ii) reports every analytical parameter of every analysis in a dedicated table, so that each result can be reproduced from the same corpus; (iii) triangulates two independent mapping algorithms modularity-based network clustering in VOSviewer and correspondence analysis in Biblioshiny so that convergence and divergence between cluster solutions can be examined rather than assumed; and (iv) reads the resulting map critically rather than descriptively, identifying the field's ChatGPT-centrism, its collaboration deficit and the early signs of thematic saturation. Table 1 positions the present study against the bibliometric reviews closest to it in scope.

Table 1. Positioning of the present study relative to prior bibliometric reviews on artificial intelligence and generative artificial intelligence in education

Study	Domain focus	Educational level	Distinctive emphasis	Difference from the present study
Zawacki-Richter et al. (2019)	Artificial intelligence (general)	Higher education	Systematic review of AI applications	Pre-dates generative AI; narrative rather than science mapping
Baek and Doleck (2020)	Artificial intelligence in education	All levels	Single-journal bibliometric profile	Restricted to one journal and to 2015–2019
Bozkurt (2023)	Generative AI, conversational agents, chatbots	All levels	Systematic review combined with bibliometrics	Not specific to higher education; earlier time window
Kavitha et al. (2024)	Artificial intelligence (general)	Higher education	Bibliometric performance analysis	Does not isolate the generative variant of AI
Flores-Velásquez et al. (2024)	Acceptance of AI technologies	Higher education	Acceptance-focused bibliometric review	Confined to technology acceptance; AI in general
Nugroho (2024)	ChatGPT	Education (general)	Trend analysis of ChatGPT research	Single-tool corpus; not higher-education-specific
Dúo-Terrón (2024)	Generative artificial intelligence	Education (general)	Analysis of scientific production	Not higher-education-specific
Ng and Ho (2025)	Generative AI in education	All levels	Mapping of the research landscape	Not higher-education-specific
Prayuda et al. (2026)	Artificial intelligence in education	All levels	Two-decade trend analysis (2000–2025)	Generative AI occupies only the final years of the window
Guo et al. (2026)	Generative AI in education	All levels	Scientometric review, 2021–2025	Not higher-education-specific
Dai et al. (2026)	Generative AI	Higher education	Emerging trends, power dynamics, global landscape	Analytical parameters not reported; single mapping tool
Present study	Generative artificial intelligence	Higher education	Dual-database, parameter-transparent, algorithmically triangulated map with a critical reading	-

Alongside the international literature, a growing body of national research supplies context-specific evidence that complements the global picture. In the Turkish context, Metli (2023) bibliometrically examined the intersection of education and artificial intelligence, while Tonbuloğlu and Avcı Akbel (2023) mapped research trends in emergency remote teaching, an adjacent field that shaped the digital infrastructure into which generative artificial intelligence tools were subsequently introduced. More directly, Büyükeke (2025) modelled the acceptance of generative artificial intelligence tools in programming courses among 305 Turkish undergraduates, using a technology acceptance model extended with self-efficacy, perceived trust, perceived risk and dependency concern and reported that perceived usefulness and perceived ease of use significantly predicted intention to use, with perceived trust operating on both. National evidence of this kind is analytically relevant to the present study for two reasons. It demonstrates that the technology-acceptance mechanisms identified in the international literature also operate in non-Anglophone higher education systems and it provides an interpretive frame for the country-level production findings reported in Section 3.4, where Türkiye appears among the moderately productive countries.

Bibliometric analysis represents an important method for systematically examining the scientific literature. Through this approach, large datasets can be analyzed to map the overall structure of a research field (Tonbuloğlu and Avcı Akbel, 2023). Bibliometric studies in the domain of generative artificial intelligence and higher education contribute to identifying influential researchers and collaboration networks (Baek and Doleck, 2020; Muskhir et al., 2024). Furthermore, keyword analyses and thematic investigations enable the identification of research trends (Muskhir et al., 2024).

Against this background, the study makes four contributions that are distinct from those of prior reviews. Methodologically, it is the first map of this specific field to combine dual-database construction with complete parameter disclosure (Table 2) and with algorithmic triangulation across two independent clustering procedures, making the reported structure verifiable rather than merely asserted. Empirically, it quantifies features of the field that descriptive maps leave implicit: the deceleration of growth beneath continuing volume expansion, the concentration of 8.75% of all citations in ten documents, the gap between a collaboration index of 2.83 and an international co-authorship rate of 10.11% and the divergence between author keywords and Keywords Plus as an indicator of terminological immaturity. Analytically, it explains rather than reports the difference between network-based and factorial cluster solutions and it shows that the two four-cluster structures they yield are complementary cross-sections of the field rather than competing accounts of it a comparison that also brings to light a substantial health-professions education subfield not previously reported in this literature. Critically, it moves beyond description to identify three structural risks facing the field ChatGPT-centrism, a collaboration deficit and early thematic saturation that carry direct implications for research funding and for institutional policy.

The aim of this study is to systematically map the scientific production dynamics, intellectual and conceptual structure and thematic evolution of generative artificial intelligence research in the context of higher education through bibliometric analysis methods. In line with this objective, the study seeks to answer the following research questions:

RQ1. What is the scientific production structure of generative artificial intelligence research in higher education?

RQ2. Around which research clusters is the intellectual structure of the generative artificial intelligence literature in higher education organized?

RQ3. Through which thematic relationships is the conceptual structure of this field of research formed?

RQ4. How are the prominent research themes in the literature on generative artificial intelligence in higher education positioned in terms of centrality and degree of development?

RQ5. How have the themes in generative artificial intelligence research in higher education evolved over time?

2. Materials and Methods

2.1. Research Design

This study adopts a bibliometric research design to examine the scientific production structure, intellectual organization, conceptual structure and thematic evolution of generative artificial intelligence research in higher education. Bibliometric analysis is a quantitative approach based on the statistical analysis of bibliographic data and is widely used to reveal the structural characteristics of research fields.

2.2. Data Sources

The data were obtained from two widely used databases due to their comprehensive indexing and academic quality standards: the Web of Science Core Collection (WoS) and Scopus.

2.3. Search Strategy and Justification of the Search Vocabulary

The literature search was conducted to identify studies examining the use of generative artificial intelligence technologies in the context of higher education. The search strategy was based on two main sets of concepts:

- Generative artificial intelligence: “generative artificial intelligence,” “generative AI,” “ChatGPT,” “large language model*,” “LLM*,” “AI chatbot*,” “GPT-4,” “GPT model*”
- Higher education context: “higher education,” “universit*,” “tertiary education”

The use of overly general and broad terms (e.g., education, learning) in isolation was avoided in order to ensure a more focused search scope. Terms within each concept set were combined using the OR operator, while

the two concept sets were combined using the AND operator.

The scope of this search vocabulary requires explicit justification. The two concept sets were deliberately constructed from umbrella terms together with the single tool name that achieved lexical dominance during the period covered, rather than from an enumeration of individual model names. Brand-specific model names (e.g., Gemini, Claude, Llama, Copilot, DeepSeek) were not entered as separate query terms, for three reasons. First, within the 2022–2025 window ChatGPT is the only generative model whose name entered the academic lexicon as a topic descriptor in its own right; studies of other models overwhelmingly index themselves under the umbrella terms already present in the query 'generative artificial intelligence', 'generative AI', 'large language model*' or 'LLM*' and are therefore captured through those terms rather than lost. Second, brand names are highly volatile: a query built around the model landscape of one year becomes unreproducible in the next, whereas umbrella terms remain stable and permit replication. Third, several model names are lexically ambiguous in database queries 'Claude' and 'Gemini', for example, retrieve substantial unrelated literature as a personal name and as an astronomical and biomedical term respectively which would have introduced a loss of precision disproportionate to any gain in recall.

Notwithstanding this rationale, a residual set of records that name only a specific model in title, abstract and keywords, without using any umbrella term, cannot be excluded. This decision is therefore also stated explicitly as a limitation in Section 5 and a model-name-inclusive sensitivity search is recommended there as a direction for subsequent replications.

Web of Science query: TS = (("generative artificial intelligence" OR "generative AI" OR ChatGPT OR "large language model*" OR LLM* OR "AI chatbot*" OR "GPT-4" OR "GPT model*") AND ("higher education" OR universit* OR "tertiary education"))

Scopus query: TITLE-ABS-KEY (("generative artificial intelligence" OR "generative AI" OR ChatGPT OR "large language model*" OR LLM* OR "AI chatbot*" OR "GPT-4" OR "GPT model*") AND ("higher education" OR universit* OR "tertiary education"))

The search was conducted on March 10, 2026. The studies were limited to the period between 2022 and 2025. Only English-language publications of the types "article" and "review" were included. In the WoS database, the SSCI, ESCI and SCIE indexes were considered. As a result of this process, 1,248 records were retrieved from WoS and 2,840 records from Scopus. One aspect of this procedure requires explicit statement. The document-type, language and index filters were applied within the database interfaces at the point of retrieval, whereas the 2022–2025 publication-year restriction was applied after retrieval, during corpus construction, rather than as a query filter. This is reported because it has a visible consequence: the retrieved set contained 142 records carrying a 2026 publication year, which were removed at the corpus-construction stage as described in Section 2.4.

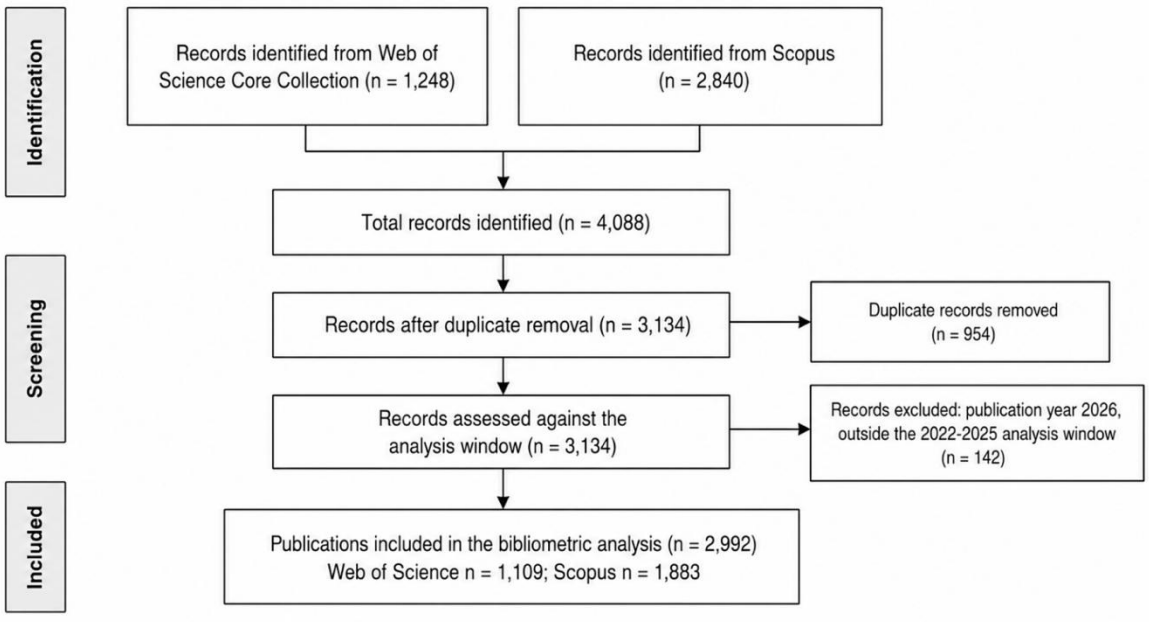


Figure 1. PRISMA 2020 flow diagram of the literature search and corpus construction process (Page et al., 2021). Records were identified in the Web of Science Core Collection (n = 1,248) and Scopus (n = 2,840) on 10 March 2026, giving 4,088 records in total. Deduplication by DOI and normalised title matching removed 954 records, all on the Scopus side, leaving 3,134 unique records. A further 142 records carrying a 2026 publication year were removed as falling outside the 2022–2025 analysis window, yielding the final analytical corpus of n = 2,992 articles and reviews (Web of Science, n = 1,109; Scopus, n = 1,883).

2.4. Screening, Selection and Eligibility Criteria

The literature screening and selection process was conducted in accordance with the PRISMA 2020 guidelines (Page et al., 2021) and is reported in full in Figure 1. A total of 4,088 records was identified: 1,248 from the Web of Science Core Collection and 2,840 from Scopus. Deduplication removed 954 records through a two-stage procedure automated matching on DOI, followed by normalised title matching (with case, punctuation and whitespace standardised) for records lacking a DOI with every automatically flagged pair verified manually before deletion. Because the Web of Science export was treated as the primary source in the merge, all duplicate removals fell on the Scopus side, leaving 3,134 unique records (Web of Science, n = 1,248; Scopus, n = 1,886). Of these, 142 records carried a 2026 publication year and therefore fell outside the analysis window; their removal yielded the final analytical corpus of 2,992 publications (Web of Science, n = 1,109; Scopus, n = 1,883), all in English and all of document type article or review.

Eligibility was operationalised at the point of retrieval rather than through a subsequent manual title-and-abstract screening stage and this is stated explicitly because it determines how the corpus should be understood. The inclusion criteria were enforced by the search query itself and by the database interface filters: the two concept sets described in Section 2.3 restricted the topic; the document-type filter restricted the set to articles and reviews; the language filter restricted it to English; and the index filter restricted the Web of Science component to the SSCI, ESCI and SCIE. The only post-retrieval reductions were the two reported above, deduplication and the publication-year restriction. No records were removed on the basis of a discretionary relevance judgement.

This is a deliberate design decision rather than an omission and it carries a specific advantage and a specific cost. The advantage is reproducibility: because no discretionary screening was applied, the corpus is fully determined by the queries, filters and retrieval date reported in Section 2.3 and any reader who applies them will obtain the same set of records which is not the case for bibliometric studies in which an unspecified number of records is removed by manual judgement. The cost is precision: records retrieved through the truncation 'universit*' in which a university appears only as the authors' affiliation rather than as the study setting and records in which generative artificial intelligence is mentioned only incidentally, remain in the corpus. This trade-off is appropriate to bibliometric mapping, whose object is a literature as it is indexed rather than a set of studies appraised for evidence synthesis, but it is a genuine constraint on the resulting map and is stated as such in Section 5.

As a single-authored study, all decisions concerning the search strategy, the filters and the deduplication procedure were made by one researcher. No inter-rater reliability coefficient is reported, because the design contains no rater-dependent screening stage for such a coefficient to describe. The criteria were specified in writing before retrieval; the cleaning and deduplication procedure was implemented as a single documented script, described in Section 2.5; and records flagged as potential duplicates were exported to a separate file and verified manually before deletion. Every reduction from the 4,088 records identified to the 2,992 analysed is therefore traceable and auditable.

Inclusion and Exclusion Criteria

Inclusion criteria;

- Addressing generative artificial intelligence within the context of higher education
- Being a peer-reviewed article or review
- Published between 2022 and 2025

Exclusion criteria;

- Studies not directly related to the research topic
- Non-academic documents such as editorials and conference abstracts
- Duplicate records

2.5. Data Collection, Cleaning and Corpus Construction

The data were exported from the WoS and Scopus databases. WoS data were obtained in .txt format, while Scopus data were retrieved in .csv format. Prior to merging, the dataset consisted of 4,088 records.

The dataset includes the following information:

- Authors
- Affiliations
- Publication year
- Journal title
- Abstract
- Keywords
- References
- Citation counts

2.5.1. Data cleaning and dataset construction

The data cleaning process was conducted systematically. Of the 4,088 records retrieved from the Web of Science and Scopus databases, 954 were removed as duplicates on the basis of DOI and normalised title matching, leaving 3,134 unique records; a further 142 records carrying a 2026 publication year were removed as falling outside the 2022–2025 analysis window. The final dataset therefore comprised 2,992 publications. Inconsistencies in data formats and fields were additionally corrected and keywords were standardised as described below.

Data cleaning was implemented in R through a single documented script applied to the merged corpus before any analysis, so that both software environments received identical input. The procedure comprised five operations. First, invisible and zero-width characters were stripped and whitespace normalised across all text fields. Second, DOIs were standardised to a bare lower-case form, with 'doi:' prefixes and the https://doi.org/ and http://dx.doi.org/ resolvers removed, so that records differing only in the presentation of their DOI could not survive as separate entries. Third, titles were

cleared of appended translated titles, which a number of records carry in the form '; [translated title]'; all such records were additionally written to a separate file for manual inspection. Fourth, affiliation strings were upper-cased and standard abbreviations expanded (UNIV → UNIVERSITY, INST → INSTITUTE, SCH → SCHOOL, DEPT → DEPARTMENT) in order to reduce the fragmentation of institutional counts. Fifth, the cited-reference field was upper-cased and duplicated 'DOI DOI' tokens collapsed. Records flagged as potential duplicates by the automated procedure were exported to a separate file and verified manually before deletion.

Keyword harmonisation warrants separate description, since it is the single cleaning decision with the greatest bearing on the conceptual and thematic results reported in Section 3. Both keyword fields and the merged keyword field derived from them, were processed by the same routine: keywords were split on the semicolon delimiter, case-normalised, cleared of invisible characters and keywords repeated within a single record were removed. A synonym dictionary was then applied to consolidate notational and morphological variants of the field's core concepts. Verified against the analysed corpus, the rules took effect as follows: 'generative artificial intelligence', 'generative artificial intelligence (GenAI)', 'generative AI (GenAI)' and 'GenAI' were consolidated into generative AI (537 keyword instances); 'large language model', 'large language model (LLM)', 'large language models (LLMs)', 'LLM' and 'LLMs' into large language models (127 instances); 'chatbot', 'AI chatbot' and 'AI chatbots' into chatbots (116 instances); 'artificial-intelligence' into artificial intelligence; and 'tertiary education' into higher education (11 instances). After harmonisation the corpus contained 6,030 unique author keywords and 2,222 unique index keywords across the 2,992 analysed documents.

Two residual features of the harmonised vocabulary are reported here rather than left implicit, because both bear on the interpretation of Figures 9 to 11. First, the consolidation of 'tertiary education' into 'higher education' means that the higher education node in those figures represents both terms. Second, two abbreviation variants were not consolidated: 'AI' (168 instances) and 'artificial intelligence (AI)' (96 instances) remain distinct from 'artificial intelligence' (761 instances) in the analysed data. The frequency of artificial intelligence as a concept is therefore distributed across three surface forms and its position in the conceptual and thematic maps is correspondingly conservative: consolidating the three would raise its frequency to 1,025 and would strengthen rather than weaken its centrality, so the reported structure is not an artefact of this residual fragmentation. No stemming was applied beyond the explicit rules listed above and no terms were merged on the basis of automatic similarity matching, so that every substitution performed on the data is recoverable from this list. This harmonisation governs the author-keyword analyses reported in Figures 9 to 11. Figure 8 is the single deliberate exception. That network was constructed on the merged keyword field, in which author keywords and the index keywords assigned by the source databases are combined and retained in the form in which the databases record them. The purpose of retaining them is analytical rather than incidental: index keywords encode how the literature is classified by the databases themselves and preserving them allows the indexed structure of the field to be compared with the vocabulary its authors choose, which is what Sections 2.8 and 3.8 set out to do. The cost of this decision is that variant forms of the same term appear as separate nodes in Figure 8 and that database descriptors which are indexing conventions rather than research themes are present in the network; both consequences are stated in Section 5. No thesaurus file was applied within VOSviewer, so the vocabulary of Figure 8 is exactly the indexed vocabulary of the corpus.

2.6. Bibliometric Analysis Tools and Analytical Parameters

The bibliometric analyses were conducted in two complementary software environments, with each analysis assigned to the environment best suited to it. Performance (descriptive) analysis, conceptual structure mapping, strategic (thematic) mapping and thematic evolution analysis were performed in the Bibliometrix R package through its interactive Biblioshiny interface (Aria and Cuccurullo, 2017), which implements the co-word and strategic-diagram procedures of Callon et al. (1991) and Cobo et al. (2011). The keyword co-occurrence network presented in Figure 8 was constructed independently in VOSviewer (van Eck and Waltman, 2010), whose unified mapping-and-clustering algorithm (Waltman et al., 2010) derives its solution from different principles to those underlying the Biblioshiny procedures, allowing the two cluster solutions to be compared rather than assumed to coincide. The two environments were supplied with deliberately different keyword vocabularies harmonised author keywords for the Biblioshiny analyses, indexed merged keywords for the VOSviewer network for the reasons set out in Section 2.5. Reporting followed the guidance on bibliometric method transparency set out by Zupic and Čater (2015) and Donthu et al. (2021).

Author keywords rather than Keywords Plus were selected as the data field for all conceptual and thematic analyses. The rationale is empirical and is set out with the corresponding evidence in Section 3.1: in this corpus author keywords outnumber Keywords Plus by a factor of 2.7 and are assigned at almost three times the per-document rate, because Keywords Plus is generated algorithmically from the titles of cited references and is therefore sparse in a field whose cited literature has not yet stabilised. Author keywords consequently provide the denser and more representative vocabulary for mapping a research front of this age. The single exception is the VOSviewer co-occurrence network in Figure 8, which was deliberately run on the merged keyword set in

order to test whether the cluster structure is robust to the choice of data field. The complete set of parameters governing every analysis reported in this study is presented in Table 2, so that each result in Section 3 can be reproduced from the same corpus.

Descriptive analysis;

- Number of publications by year
- Most productive authors
- Country-based production
- Citation indicators

Network analysis;

Relationships between concepts were examined through keyword co-occurrence analysis.

Conceptual structure analysis;

The conceptual structure of the field was identified by analyzing the clustering of keywords.

Thematic analysis;

Themes were classified according to their centrality and density values:

- Motor themes
- Basic themes
- Niche themes
- Emerging themes

Additionally, thematic evolution analysis was conducted to examine changes over time.

Table 2. Analytical parameters of the bibliometric analyses reported in this study

Analysis (output)	Software	Data field	Unit selection / threshold	Normalisation / weighting	Clustering / extraction method	Additional settings
Study selection (Figure 1)	-	Full records	4,088 records screened; 2,992 retained	-	-	PRISMA 2020 (Page et al., 2021); DOI then normalised-title deduplication
Performance and descriptive indicators (Figures 2–7)	Bibliometrix / Biblioshiny	Full records	All 2,992 documents	-	-	Full counting; country attribution by author affiliation; citation counts as indexed on 10 March 2026
Keyword co-occurrence network (Figure 8)	Network matrix built in Bibliometrix, exported in Pajek format, laid out and clustered in VOSviewer	Merged keyword field (author keywords + database index keywords), retained as indexed (Section 2.5)	50 most frequent keywords; effective minimum occurrence = 111; full counting	Association strength	Modularity-based clustering in VOSviewer; minimum cluster size = 1; 4 clusters obtained	Resulting network: 50 nodes, 646 edges; isolates removed; node size proportional to occurrences, edge width to link strength, colour to cluster
Conceptual structure map (Figure 9)	Bibliometrix / Biblioshiny	Author keywords, harmonised in R prior to analysis (Section 2.5)	50 most frequent terms	-	Multiple Correspondence Analysis (MCA); terms partitioned by k-means, k = 4	Terms projected onto the first two factorial dimensions; proximity denotes similarity of co-occurrence profiles
Strategic (thematic) map (Figure 10)	Bibliometrix / Biblioshiny	Author's Keywords, harmonised in R prior to analysis (Section 2.5)	250 terms; minimum cluster frequency = 5 per thousand documents	Equivalence (inclusion) index	Louvain community detection (Blondel et al., 2008)	Quadrant boundaries set at the median of observed centrality and density values; no stop-word or in-application synonym list applied
Thematic evolution (Figure 11)	Bibliometrix / Biblioshiny	Author's Keywords, harmonised in R prior to analysis (Section 2.5)	250 terms; minimum cluster frequency = 5 per thousand documents	Inclusion index weighted by word occurrences	Louvain community detection (Blondel et al., 2008)	Cutting points at 2023 and 2024; flow width proportional to the weighted inclusion index

2.7. Strategic (Thematic) Map: Computation of Centrality and Density

Themes were positioned on a strategic diagram using the co-word procedure formalised by Callon et al. (1991) and

operationalised in Biblioshiny by Aria and Cuccurullo (2017). Author keywords were first normalised by the equivalence index, e(ij) = c(ij)^2 / (c(i) x c(j)), where c(ij) is the number of documents in which keywords i and j co-occur and c(i) and c(j) are the numbers of documents in which each keyword occurs separately. The resulting weighted co-word network was partitioned into clusters by the Louvain algorithm (Blondel et al., 2008), which detects communities by greedy modularity optimisation, iteratively reassigning terms between communities and aggregating them until modularity ceases to improve.

Two indices were then computed for each cluster. Centrality, which measures the strength of a theme's links to other themes and therefore its importance to the field as a whole, was calculated as c = 10 x SUM e(kh), where k is a keyword belonging to the cluster under consideration and h is a keyword belonging to any other cluster. Density, which measures the internal cohesion of a theme and therefore its degree of development, was calculated as d = 100 x (SUM e(ij)) / w, where i and j are keywords belonging to the same cluster and w is the number of keywords in that cluster.

Clusters were assigned to quadrants by splitting each axis at the median of the observed centrality and density values across all retained clusters, following Cobo et al. (2011). No fixed absolute cut-off value was applied. This is a deliberate methodological choice rather than an omission: centrality and density are scale-dependent on corpus size and keyword frequency, so an absolute threshold would be neither meaningful within a single corpus nor comparable across corpora, whereas a median split renders the classification internal to the corpus analysed and therefore interpretable. The four resulting quadrants are motor themes (high centrality, high density), basic and transversal themes (high centrality, low density), niche themes (low centrality, high density) and emerging or declining themes (low centrality, low density). It follows that a theme's position in Figure 10 should be read as its standing relative to the other themes in this corpus, not as an absolute measure of its maturity.

Thematic evolution was analysed by dividing the corpus into three time slices at the cutting points 2023 and 2024, computing the strategic map within each slice using identical parameters and linking themes across adjacent slices by the inclusion index weighted by word occurrences. The width of each flow in Figure 11 is proportional to that index.

2.8. Comparability of the Network and Factorial Cluster Solutions

Because the co-occurrence network (Figure 8) and the conceptual structure map (Figure 9) are produced by different algorithms operating on deliberately different inputs, the two solutions are reported and interpreted separately rather than merged. Figure 8 is a network partition: keywords drawn from the merged, indexed keyword field are clustered in the full high-dimensional co-occurrence space by optimising a modularity-based quality function. Figure 9 is a factorial projection: correspondence analysis reduces the harmonised author-keyword structure to a small number of latent dimensions and terms are then partitioned within that reduced space.

Both procedures return four clusters, but they are not the same four and this is the point of running them together rather than a defect in either. Because Figure 8 operates on indexed vocabulary, its clusters track the disciplinary settings in which the research is carried out; because Figure 9 operates on harmonised author keywords, its clusters track the conceptual concerns that authors themselves name. The comparison is therefore between how this literature is classified and how it describes itself and neither view is reducible to the other. The substantive relationship between the two solutions is set out in Section 3.8.

2.9. Data Visualization and Figure Preparation

Various bibliometric visualization techniques were employed to present the research findings in a clear and comprehensible manner. Time-series graphs were generated to illustrate trends in scientific production, while country production maps were used to reveal the geographical distribution of research outputs. Furthermore, keyword co-occurrence networks were constructed to visualize conceptual relationships within the literature. Conceptual structure maps and thematic maps were produced using Biblioshiny and network-based visualizations were generated in high resolution using the VOSviewer software.

For this version the figures have been re-prepared with legibility in mind, following the reviewers' comments. The keyword co-occurrence network in Figure 8, identified in review as the least legible, has been re-exported from VOSviewer at the maximum available resolution, with enlarged label fonts and increased label-to-node scaling; it is reproduced here at approximately 245 dpi, against 111 dpi in the previous version. The PRISMA diagram in Figure 1 has been redrawn to the corrected selection counts reported in Section 2.4 and the thematic evolution diagram in Figure 11 has been regenerated in Biblioshiny using the parameters given in Table 2. Vector versions (SVG or PDF) of the figures produced in Biblioshiny and VOSviewer can be supplied to the publisher at the production stage should the journal's typesetting workflow require them.

3. Results

This section presents the bibliometric characteristics of generative artificial intelligence research in higher education. The analyses encompass the scientific production structure, temporal trends, distributions of authors and countries, citation impact and the conceptual structure of the field.

3.1. Scientific Production Structure of the Field

The main indicators of the bibliometric dataset analyzed in this study are presented in Figure 2. The dataset covers the period between 2022 and 2025 and consists

of 2,992 documents obtained from 728 different sources. During the examined period, the average annual scientific production was approximately 748 documents. The average number of citations per document was calculated as 18.46, indicating that the field demonstrates a considerable level of academic impact.

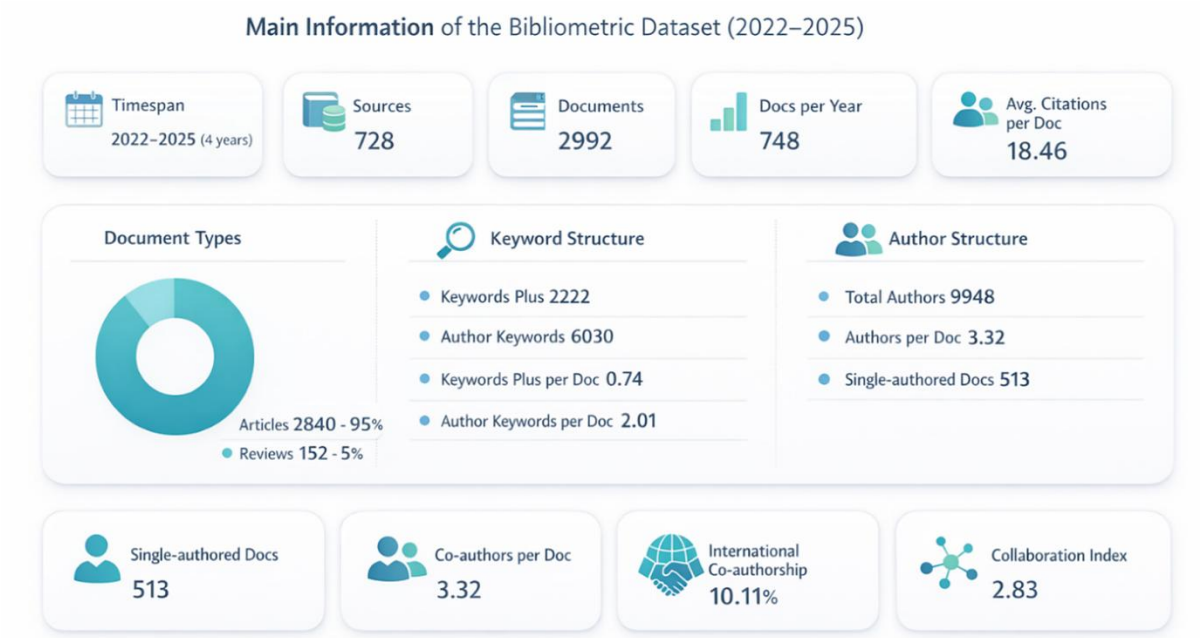


Figure 2. Main characteristics and descriptive statistics of the bibliometric corpus (2022–2025). Generated in Biblioshiny from the merged Web of Science–Scopus corpus (n = 2,992 documents; 728 sources; 9,948 authors; mean 18.46 citations per document). Note that the Keywords Plus count reported in the figure is 2,222 terms, which is consistent with the ratio of 0.74 Keywords Plus per document also shown.

The analysis of document types indicates that the dataset is predominantly composed of research articles (n = 2,840; 94.92%), while review studies constitute a relatively small proportion (n = 152; 5.08%). This distribution suggests that the field is largely shaped by original and empirical research and is undergoing a dynamic development process.

Findings related to the keyword structure reveal the conceptual profile of the field. A total of 2,222 Keywords Plus and 6,030 author keywords were identified in the analysed corpus, corresponding to 0.74 Keywords Plus and 2.01 author keywords per document. The contrast between these two figures is analytically meaningful rather than incidental. Author keywords outnumber Keywords Plus by a factor of 2.7 and are assigned at almost three times the per-document rate, which is the reverse of the pattern typically observed in mature fields. Keywords Plus is generated algorithmically from the titles of the references a document cites and therefore reflects the vocabulary of the established literature on which a field draws; author keywords reflect the vocabulary that authors themselves choose. In a field as young as this one, the cited literature has not yet stabilised into a shared indexing vocabulary, which depresses Keywords Plus, while authors coin and apply terminology freely, which inflates author keywords. The divergence is thus a quantitative signature of terminological immaturity: the field is conceptually broad but has not yet converged on a standard descriptive vocabulary. This finding also provides the empirical justification for the methodological decision reported in Section 2.6, namely that author keywords rather than Keywords Plus were used as the data field for the conceptual and thematic analyses that follow. An examination of the authorship structure shows that a total of 9,948 authors contributed to the dataset, with an average of 3.32 authors per document. Single-authored studies are relatively limited in number (n = 513). This finding demonstrates that the field is largely characterized by a collaborative research structure. The rate of international co-authorship was calculated as 10.11%, while the collaboration index was 2.83. These results indicate a moderate level of international collaboration in the field, with multi-authored studies being predominant. These two indicators should be read together, because they point in opposite directions. A collaboration index of 2.83 and a mean of 3.32 authors per document, against only 513 single-authored studies (17.1% of the corpus), indicate that co-authorship is the clear norm in this field. An international co-authorship rate of 10.11%, however, means that approximately nine out of every ten collaborative papers are produced entirely within a single country. The field is therefore collaborative but not international: teams form readily, but almost always domestically. This is a structural weakness rather than a neutral descriptive observation. The questions that dominate this literature assessment security, integrity policy, equity of access, disciplinary

norms for permissible use are precisely those whose answers vary by national higher education system and which consequently require cross-system comparison if they are to be answered well. A field expanding this rapidly while collaborating this locally risks accumulating parallel national literatures that replicate one another's findings, rather than a cumulative comparative evidence base.

3.2. Annual Scientific Production

To reveal the temporal development of scientific production in the research field, the distribution of publications by year was analyzed. The findings are presented in Figure 3.

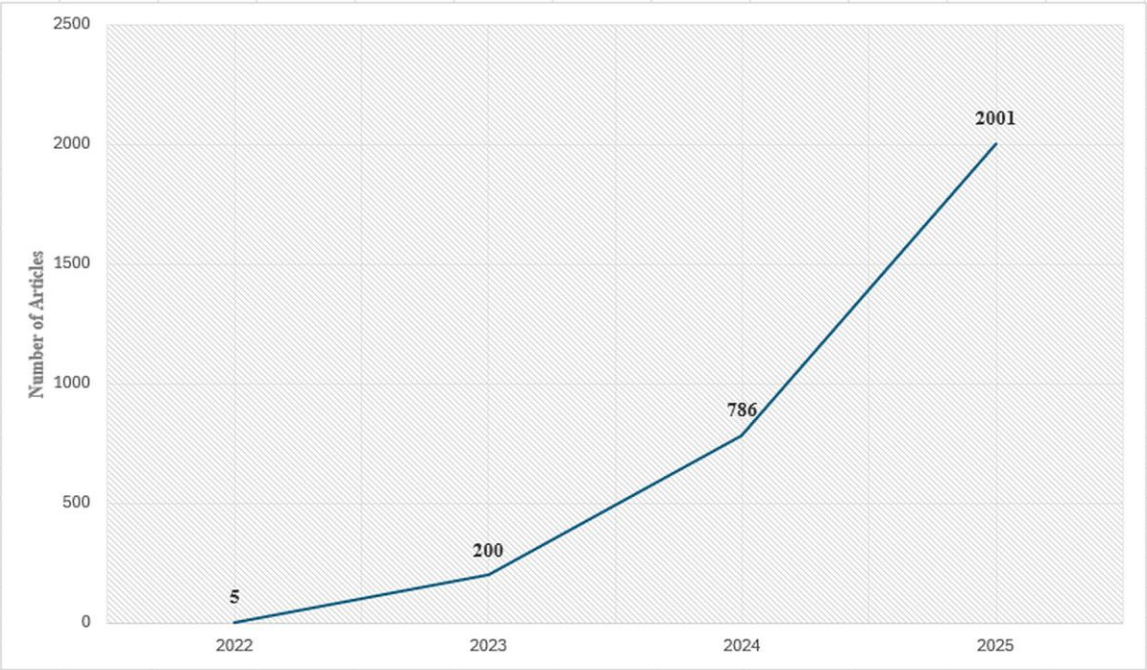


Figure 3. Annual scientific production of research on generative artificial intelligence in higher education (2022–2025). Generated in Biblioshiny; articles and reviews only. Annual document counts: 2022, n = 5; 2023, n = 200; 2024, n = 786; 2025, n = 2,001. The 2025 count should be treated as a lower bound, since records indexed after 10 March 2026 are not captured.

An examination of Figure 3 reveals that scientific production in the field of generative artificial intelligence in higher education exhibits a clear and sharp upward trend over the analyzed period. While the level of production was quite limited in 2022, with only 5 publications, this number increased significantly to 200 in 2023, indicating a remarkable rise. This upward trend accelerated further in 2024, reaching a total of 786 publications. The most striking increase occurred in 2025, when the number of publications surged to 2,001, far exceeding the levels observed in previous years.

Expressed in relative terms, the trajectory is considerably more informative than the raw counts. Output grew forty-fold between 2022 and 2023, by a further 293.0% in 2024 and by 154.6% in 2025, giving a 400-fold expansion across the four-year window as a whole. The pattern of these rates matters: growth decelerated markedly from 3,900% to 293.0% to 154.6% even as absolute volume continued to rise steeply. This combination is the characteristic signature of a research field moving off an explosive initial phase onto a steep but decelerating growth path, rather than one still accelerating without constraint. Two-thirds of the entire corpus (66.9%, n = 2,001) was published in the final year alone and 93.1% (n = 2,787) in the final two years. This concentration is double-edged. It confirms the currency of the field, but it also means that the overwhelming majority of the literature has been in circulation for less than two years and has had minimal opportunity to be cited, replicated or challenged. The findings reported below should therefore be read as a map of an active research front rather than of a consolidated body of knowledge.

These findings suggest that the research field has rapidly transitioned from an early developmental stage to a phase of maturation. The exponential growth observed particularly after 2023 appears to be closely associated with the increasing academic interest and research investment in generative artificial intelligence technologies within the context of higher education. Moreover, the high level of scientific production reached in 2025 indicates that the field is not only gaining popularity but is also entering a phase of institutionalization and becoming a central focus within the research agenda.

3.3. Most Productive Authors

To identify the most scientifically productive researchers in the field, an author productivity analysis was conducted. The findings are presented in Figure 4.

An examination of Figure 4 indicates that the most prolific author in the analyzed dataset is Chiu T. (n = 14). Chiu T. is followed by Gasevic D. (n = 12) and Bozkurt A. (n = 9). Similarly, Chan C. (n = 9) and Li X. (n = 9) are also among the authors with high levels of productivity. These findings suggest that certain researchers play a leading role in knowledge production within the field.

When authors with moderate levels of productivity are considered, Khalaf Z., Li Y., Noroozi O., Perkins M. and Roe J. each have 7 publications. This group can be regarded as researchers who contribute consistently to the development of the field and play an important role in the expansion of the literature. Overall, the distribution of author productivity indicates that the field is concentrated around a set of core researchers.

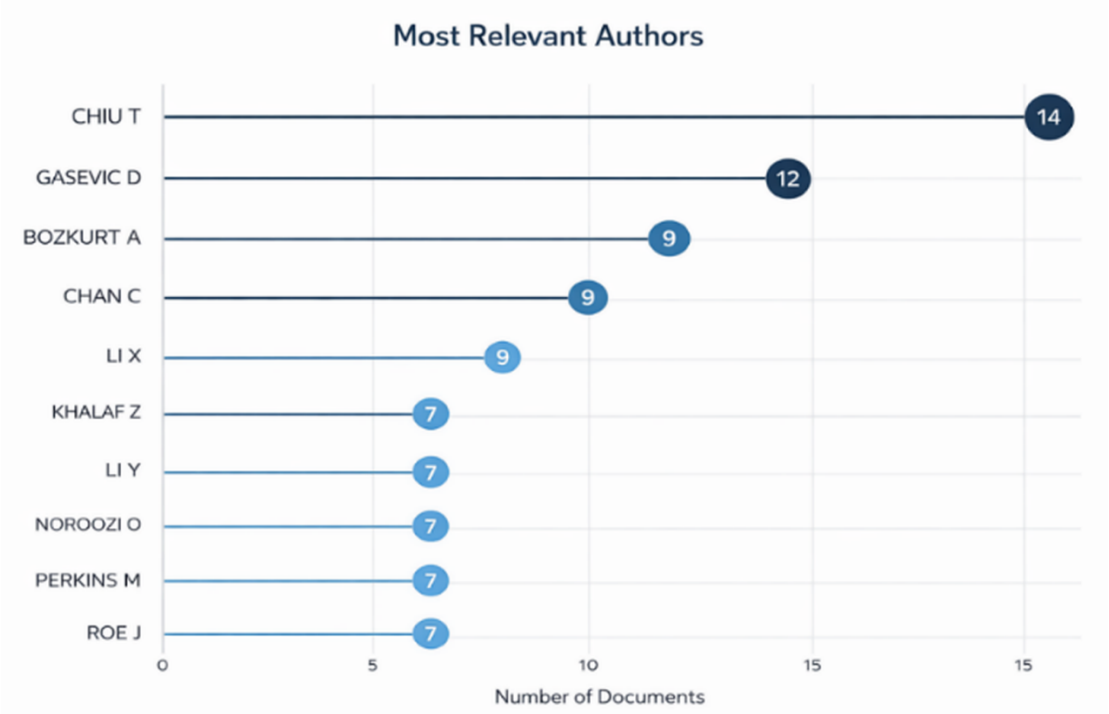


Figure 4. Most productive authors in the research field, by number of documents (full counting).Generated in Biblioshiny; author names as disambiguated in the source databases. The ten most productive authors account for 88 documents in total, or 2.9% of the corpus.

The distribution of author productivity is, however, notably flat for a field of this size and this flatness is itself a finding. The most productive author accounts for 14 of 2,992 documents (0.47%) and the ten most productive authors together account for 88 documents, or 2.9% of the corpus, against 9,948 contributing authors in total a ratio of one document per 3.3 authors. There is, in other words, no dominant research school in this field. It is being built by a very large number of researchers each contributing one or two papers, with only a thin layer of sustained specialists above them. This is consistent with a young research front with low barriers to entry and it carries a direct implication for the interpretation of the thematic results reported below: thematic structure in such a field is driven by the aggregate vocabulary of many occasional contributors rather than by the programmatic agendas of a few dominant groups, which makes the resulting map broad but comparatively shallow.

3.4. Institutional and Country Production

To reveal the geographical distribution of scientific production in the research field, country-based publication output was analyzed. The findings are presented in Figure 5.

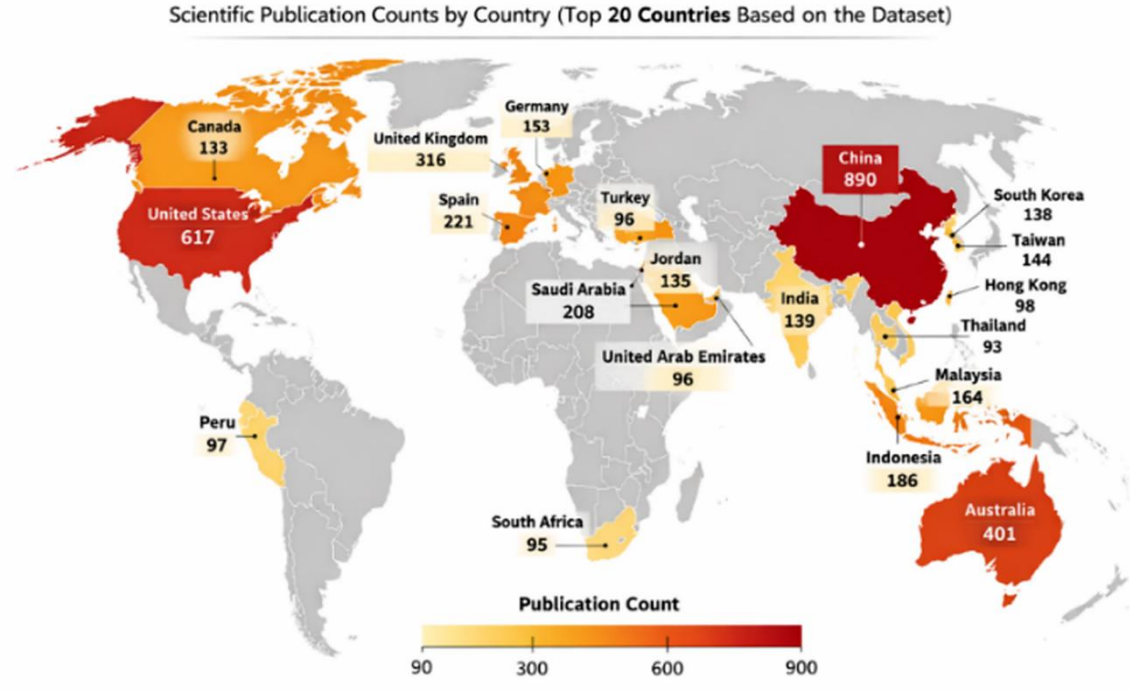


Figure 5. Country scientific production in research on generative artificial intelligence in higher education. Generated in Biblioshiny on the basis of author affiliations using full counting, so that a document authored across k countries contributes one unit to each of them; country counts consequently sum to more than the number of documents and should not be aggregated. Selected values: China, n = 890; United States, n = 617; Australia, n = 401; Türkiye, n = 96.

An examination of Figure 5 indicates that scientific production in generative artificial intelligence research in higher education is concentrated in specific countries. China ranks as the most productive country by a considerable margin, with 890 publications. It is followed by the United States (n = 617) and Australia (n = 401). These three countries constitute a significant portion of global scientific output in the field and occupy a central position in research activities.

In Europe, the United Kingdom (n = 316) and Spain (n = 221) stand out as leading contributors. Germany (n = 153) and Türkiye (n = 96) are among the countries with moderate levels of production. In the Middle East, Saudi Arabia (n = 208) and Jordan (n = 135) emerge as notable contributors at the regional level due to their relatively high publication outputs.

The Asia-Pacific region demonstrates a high level of research productivity. Indonesia (n = 186), Malaysia (n = 164), India (n = 139), South Korea (n = 138) and Taiwan (n = 144) exhibit considerable levels of output. In addition, countries such as Hong Kong (n = 98) and the Philippines (n = 93) also contribute to the development of the field. These findings indicate that the Asia-Pacific region is increasingly assuming a central role in the advancement of the field.

In the Americas, Canada (n = 133) and Peru (n = 97) are among the notable contributors, while in Africa, South Africa (n = 95) stands out as a leading country in regional production. Furthermore, countries such as the United Arab Emirates (n = 96) also make meaningful contributions to the research field. Overall, scientific production appears to be predominantly concentrated in the Asia-Pacific and North American regions, with Europe and the Middle East playing complementary and supportive roles.

Türkiye's position in this distribution warrants specific comment, since the country appears in the dataset with 96 publications. In relative terms this places Türkiye in the same output band as the United Arab Emirates (n = 96), Peru (n = 97), Hong Kong (n = 98) and South Africa (n = 95) and marginally above the Philippines (n = 93), while representing 10.8% of China's output and 15.6% of that of the United States. More instructive, however, is the comparison with higher education systems of broadly comparable scale and with comparable recent investment in digital transformation. Türkiye's output falls well below that of Saudi Arabia (n = 208), Indonesia (n = 186), Malaysia (n = 164), Taiwan (n = 144), India (n = 139) and Jordan (n = 135) and below Germany (n = 153). The comparison with Jordan is the most pointed of these, since the Jordanian higher education system is a fraction of the size of Türkiye's yet produces 40.6% more publications in this field. Türkiye therefore participates in this literature, but not yet at a level proportionate to the scale of its higher education system and the shortfall is with regional peers rather than with the global leaders. Two qualifications apply to this reading and are developed in Sections 4.3 and 5: the English-language restriction and the exclusion of nationally indexed outlets mean that the figure of 96 is a lower bound on Turkish scholarly activity in this area rather than a complete count of it.

3.5. Corresponding Authors' Countries

To examine the distribution of scientific production in the research field based on the countries of corresponding authors, a country-based analysis was conducted. The findings are presented in Figure 6.

An examination of Figure 6 shows that China has the highest level of scientific production based on the distribution of corresponding authors' countries. China is followed by the United States and the United Kingdom. These countries stand out as leading actors in the field not only in terms of total publication output but also at the level of corresponding authorship. Australia is also among the countries with a notably high level of production.

The figure further categorizes publications into single-country publications (SCP) and multiple-country publications (MCP). The findings indicate that, in many countries, publications originating from single-country collaborations (SCP) are more dominant. This suggests that research activities are largely concentrated at the national level. However, in countries with higher production capacity, multi-country collaborations (MCP) appear to be more prominent.

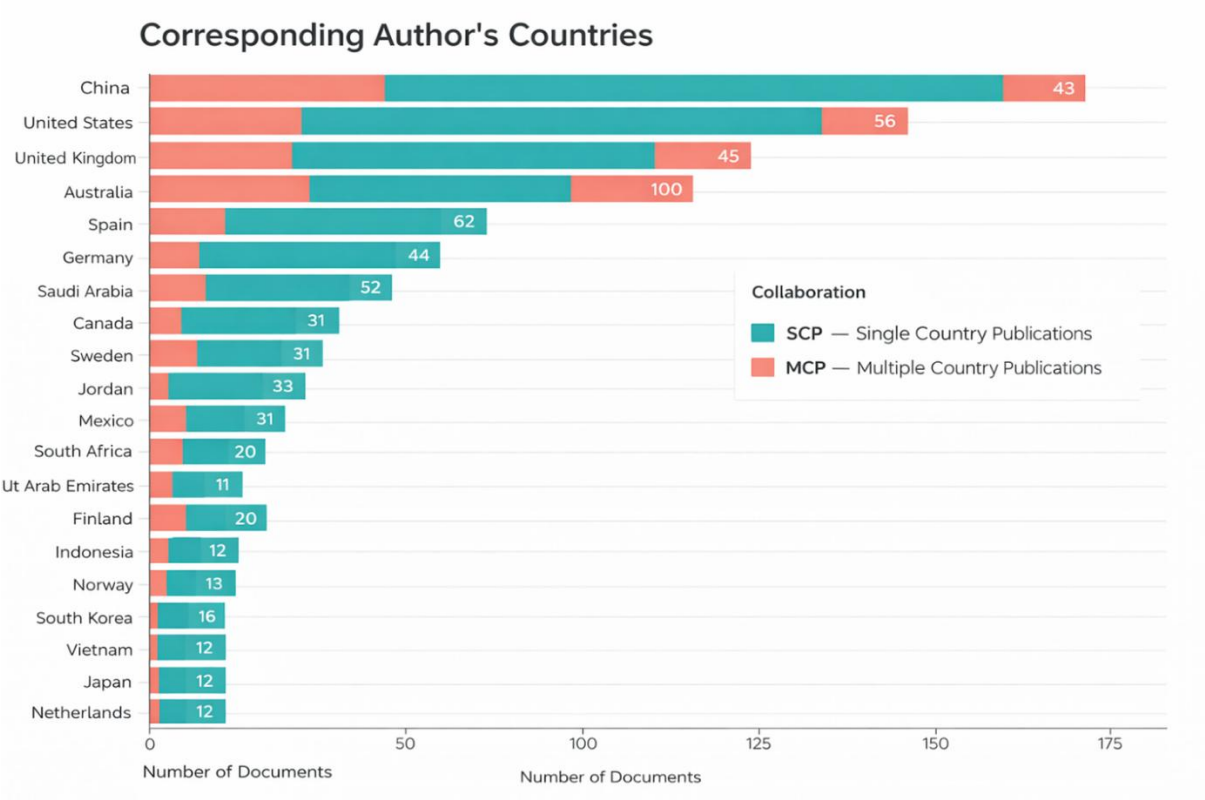


Figure 6. Distribution of publications by corresponding author's country, decomposed into single-country. Publications (SCP) and multiple-country publications (MCP). Generated in Biblioshiny; the multiple-country publication share across the corpus is 10.11%.

Among European countries, the United Kingdom, Spain and Germany are notable contributors; in the Middle East, Saudi Arabia stands out; and in North America, Canada is among the leading countries. Additionally, countries such as Sweden, Jordan and Mexico demonstrate moderate levels of production performance. Although exhibiting lower levels of production, countries including South Africa, the United Arab Emirates, Finland, Indonesia, Norway, South Korea, Vietnam, Japan and the Netherlands are also represented in the dataset. Overall, the distribution of corresponding authorship indicates that scientific production is concentrated in specific countries, which assume leading roles in research processes. Furthermore, international collaborations appear to be more prominent particularly in countries with higher production capacity.

3.6. Citation Impact

To identify the most influential studies in the research field, a citation analysis was conducted and the documents with the highest global citation counts are presented in Figure 7.

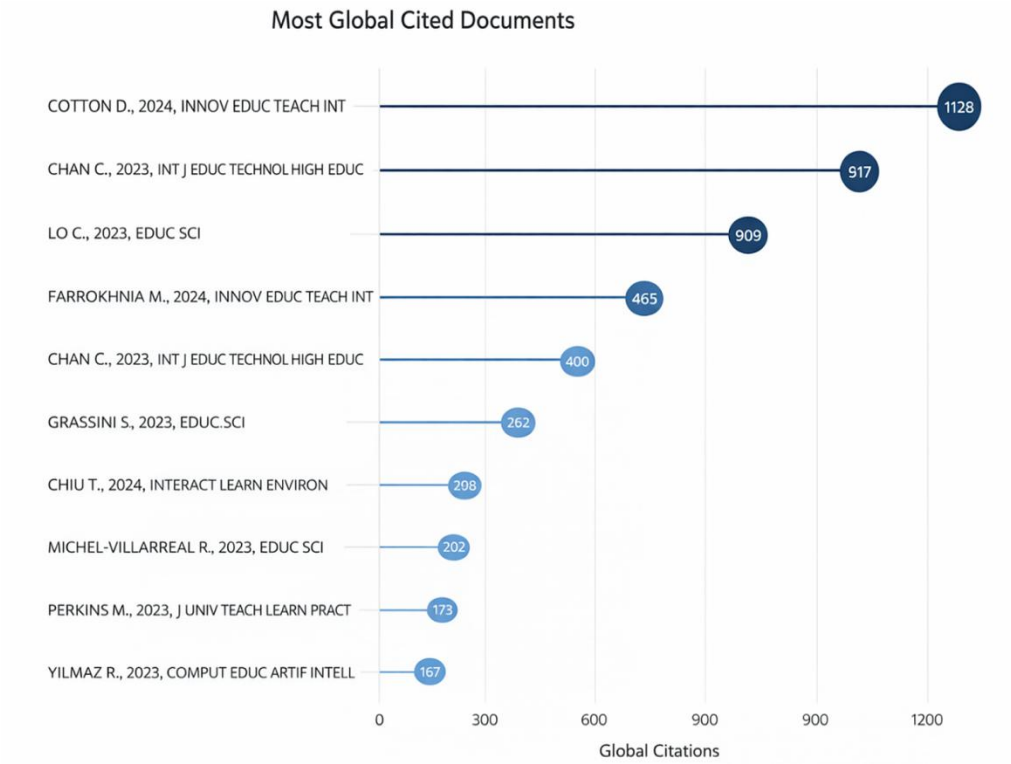


Figure 7. Ten most globally cited documents in the corpus, ranked by total global citations as indexed in the source databases on 10 March 2026. Generated in Biblioshiny. These ten documents account for 4,831 citations, equivalent to 8.75% of all citations accumulated by the corpus.

An examination of Figure 7 reveals that citation impact within the dataset is concentrated around a limited number of publications. The study with the highest number of citations is that of Cotton D. (2024), which has received 1,128 global citations. This is followed by Chan C. (2023) with 917 citations and Lo C. (2023) with 909 citations. These findings indicate that these studies possess high visibility and influence within the field.

Among studies with a moderate-to-high level of citations, the work by Farrokhnia M. (2024) (n = 465) and another study by Chan C. (2023) (n = 400) stand out. Additionally, the study published by Grassini S. (2023) has made a notable impact in the literature with 262 citations. These studies can be considered key reference points contributing to both the theoretical and applied development of the field.

Studies with lower yet still significant citation levels include those by Chiu T. (2024) (n = 208), Michel-Villarreal R. (2023) (n = 202), Perkins M. (2023) (n = 173) and Yılmaz R. (2023) (n = 167). These works play a complementary and supportive role within the expanding body of literature. Overall, the distribution of citations indicates the presence of a “high-impact core” formed by a select group of publications. This suggests that the knowledge structure of the field is shaped around a set of leading studies and that academic impact is distributed unevenly. The labels in Figure 7 identify documents contained in the analysed corpus by first author and publication year exactly as they are recorded in the source databases; they are data points of the citation analysis rather than works cited by this study and are therefore not listed in the reference list.

The degree of this concentration can be quantified precisely. With a mean of 18.46 citations per document across 2,992 documents, the corpus has accumulated approximately 55,200 citations in total. The ten documents shown in Figure 7 account for 4,831 of these that is, 8.75% of all citations in the field are generated by 0.33% of its documents, while the three most-cited documents alone (n = 2,954 citations) account for 5.4%. This is a steeply skewed distribution even by the standards of citation data and it has an identifiable cause. All three leading documents were published in 2023 or earlier and address academic integrity or the pedagogical implications of ChatGPT. In a field where two-thirds of the literature appeared only in 2025, early publication confers a decisive citation advantage and the works that happened to appear first were agenda-setting commentaries rather than empirical studies. The 'high-impact core’ identified here should therefore be understood as a first-mover effect operating on framing pieces, rather than as evidence that the field’s most methodologically substantial work has already been identified. This is a caution regarding the maturity of the citation structure, not a criticism of the works concerned.

3.7. Intellectual Structure of the Field

To reveal the intellectual structure of the research field, a keyword co-occurrence analysis was conducted. The results of the analysis are presented in Figure 8.

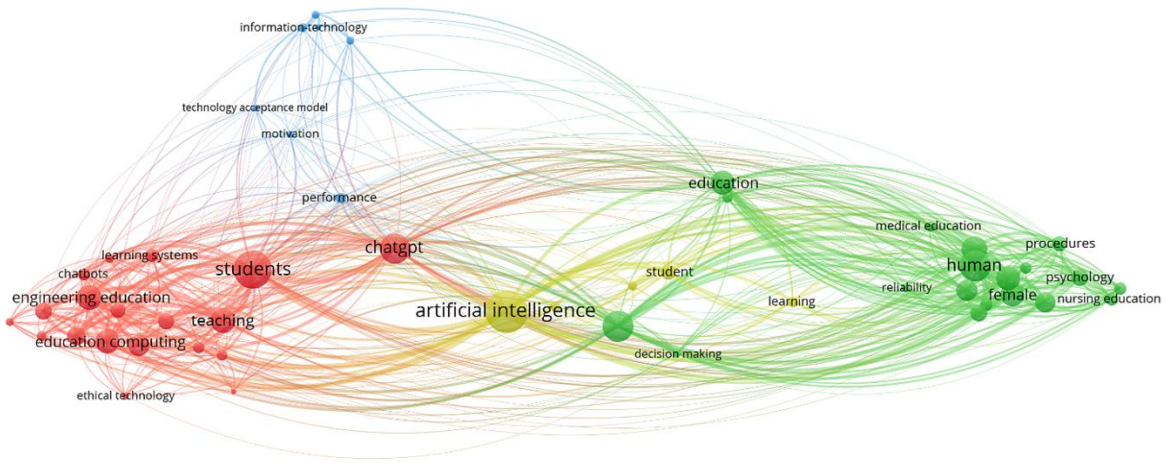


Figure 8. Keyword co-occurrence network of research on generative artificial intelligence in higher education. The network matrix was constructed in Bibliometrix from the merged keyword field author keywords combined with the index keywords assigned by the source databases, retained as indexed and exported in Pajek format; layout and clustering were performed in VOSviewer using association strength normalisation and modularity-based clustering. The network comprises the 50 most frequent keywords (effective minimum occurrence = 111) connected by 646 co-occurrence links and resolves into four clusters. Node size is proportional to occurrence frequency, edge thickness to co-occurrence link strength and colour denotes cluster membership. Because index keywords are retained, variant forms of the same term appear as separate nodes; see Sections 2.5 and 5.

The network presented in Figure 8 was constructed from the merged keyword field, in which author keywords and the index keywords assigned by the source databases are combined and retained in their indexed form. Nodes represent keywords, node size is proportional to occurrence frequency, edge thickness to co-occurrence link strength and colour denotes cluster membership. The network comprises the 50 most frequent keywords, connected by 646 co-occurrence links and resolves into four clusters. Because this analysis deliberately retains index vocabulary, it maps how the literature is classified by the databases as well as how it is described by its authors; the complementary author-defined view is reported in Section 3.8.

The first and largest cluster is organised around instructional practice and educational technology. It contains students, teaching, engineering education, education computing, curricula, learning systems, chatbots, computer aided instruction, university students and ethical technology, together with ChatGPT, which sits at the boundary between this cluster and the centre of the network. The presence of language model, large language model, generative AI and artificial intelligence tools within this same cluster is analytically significant: the technical vocabulary of generative systems has been absorbed into the instructional literature rather than remaining a separate technological domain. In this network, in other words, generative artificial intelligence does not constitute a technological cluster of its own.

The second cluster is defined by health-professions education. It brings together medical education, nursing education and nursing student with the biomedical index vocabulary that accompanies this literature in the source databases: human, humans, female, male, adult, psychology, procedures, questionnaire, surveys and questionnaires, reliability and human experiment. Two observations follow. Substantively, the size and cohesion of this cluster show that health-professions education constitutes a distinct and highly active subfield of generative artificial intelligence research in higher education a specialisation that is not visible in analyses restricted to author keywords and that is largely absent from the existing bibliometric literature on this topic. Methodologically, the demographic and study-design descriptors within this cluster are artefacts of biomedical indexing conventions rather than research themes in their own right and should be read as markers of that indexing tradition; the point is restated as a limitation in Section 5.

The third cluster is small but conceptually coherent and is organised around technology acceptance, comprising technology acceptance model, motivation, performance, acceptance and information-technology. Although it accounts for only a small share of the network's nodes, its separation from the two large clusters is meaningful: acceptance research forms a methodologically distinct tradition within this literature, drawing on its own theoretical vocabulary and connecting to the rest of the field mainly through the central nodes rather than through direct links to the applied clusters.

The fourth cluster is centred on artificial intelligence as a general term, which occupies a central position in the layout and mediates between the instructional cluster on one side and the health-professions cluster on the other. Here position matters more than size: the most general term in the field functions as the connective tissue

between two specialised communities that are otherwise only weakly linked to each other.

Read as a whole, the network indicates that this literature is stratified less by technological distinctions than by the disciplinary settings in which generative artificial intelligence is studied. The principal division runs between an instructional-technology tradition and a health-professions education tradition, with technology acceptance forming a smaller methodological enclave and general artificial intelligence vocabulary bridging them. This is a substantively different picture from the one produced by the author-keyword analysis reported in Section 3.8 and the relationship between the two is examined there.

3.8. Conceptual Structure of the Field

To reveal the conceptual structure of the research field, a conceptual structure mapping analysis based on keywords was conducted. The findings are presented in Figure 9.

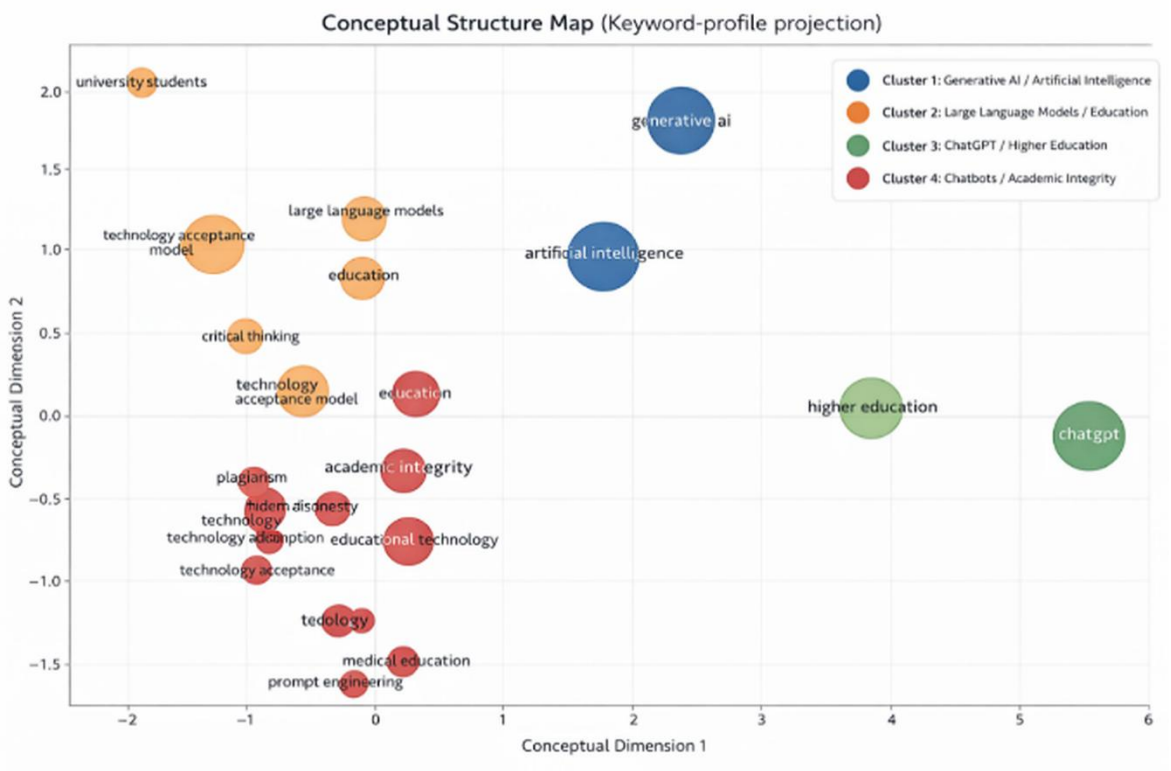


Figure 9. Conceptual structure map of the research field. Produced in Biblioshiny by multiple correspondence analysis of the 50 most frequent author keywords, with terms partitioned by k-means (k = 4) in the space defined by the first two factorial dimensions. Spatial proximity denotes similarity between keywords' co-occurrence profiles and should not be read as co-occurrence frequency. This four-cluster solution and the four-cluster solution in Figure 8 are not the same four; the two are compared in Section 3.8.

An examination of the conceptual structure map presented in Figure 9 indicates that keywords are positioned around distinct thematic clusters, reflecting the main research axes of the field. In the map, nodes represent keywords, node sizes indicate the relative importance of these concepts in the literature and spatial proximity reflects conceptual similarity and thematic relationships. The analysis reveals that the field is structured around four main thematic clusters.

The first cluster represents the technological core of the field. Within this cluster, concepts such as generative AI and artificial intelligence occupy central positions and exhibit strong connections with terms such as large language models. This suggests that the research field is largely shaped by the development and technical infrastructure of generative artificial intelligence technologies.

The second cluster focuses on the applications of generative artificial intelligence in educational contexts. In this cluster, concepts such as education, higher education, ChatGPT and chatbots are prominent. This thematic structure indicates that research on the integration of artificial intelligence tools into teaching processes and their use in educational environments occupies an important place in the literature.

The third cluster is organized around ethics, assessment and academic integrity. It includes concepts such as academic integrity, plagiarism, ethics, student assessment and technology adoption. These findings demonstrate that ethical debates and transformations in assessment processes associated with the use of generative artificial intelligence in education have become significant research foci in the literature.

The fourth cluster focuses on pedagogical approaches

and learning processes. In this cluster, concepts such as teacher education, engineering education, critical thinking, education and the technology acceptance model are prominent. This thematic area represents studies examining the effects of AI-supported learning environments on teaching methods and learning outcomes.

Overall, the conceptual structure map indicates that research on generative artificial intelligence is organized around four main axes: (i) technological development, (ii) educational applications, (iii) ethics and assessment processes and (iv) pedagogical transformation. This multilayered structure clearly demonstrates the interdisciplinary nature of the field and the strong interaction between technological innovations and educational practices.

The four clusters obtained in Figure 9 are not the same four obtained in Figure 8 and the comparison between them is one of the more informative results of this study. As explained in Section 2.8, the two figures are produced by different algorithms operating on deliberately different inputs. Figure 8, built on indexed vocabulary, divides the field by disciplinary setting: instructional technology, health-professions education, technology acceptance and a general artificial intelligence core. Figure 9, built on harmonised author keywords, divides it by conceptual concern: technological development, educational application, ethics and assessment and pedagogical transformation. The two views are complementary rather than competing. The first shows where this research is being conducted, the second what it is about; reporting only one of them would have produced a partial map. It is notable, for instance, that the ethics and assessment dimension is clearly visible in the author-keyword solution but does not form a separate cluster in the indexed solution, which suggests that academic integrity is a concern authors foreground in their own terms before database indexing has caught up with it. Conversely, the health-professions education cluster is prominent in the indexed solution yet invisible in the author-keyword solution. Triangulating the two procedures is what makes both observations available and this is the principal methodological reason for having run them in parallel.

3.9. Thematic Structure of the Literature

To examine the centrality and density levels of themes in the research field, a thematic map analysis was conducted. The results are presented in Figure 10.

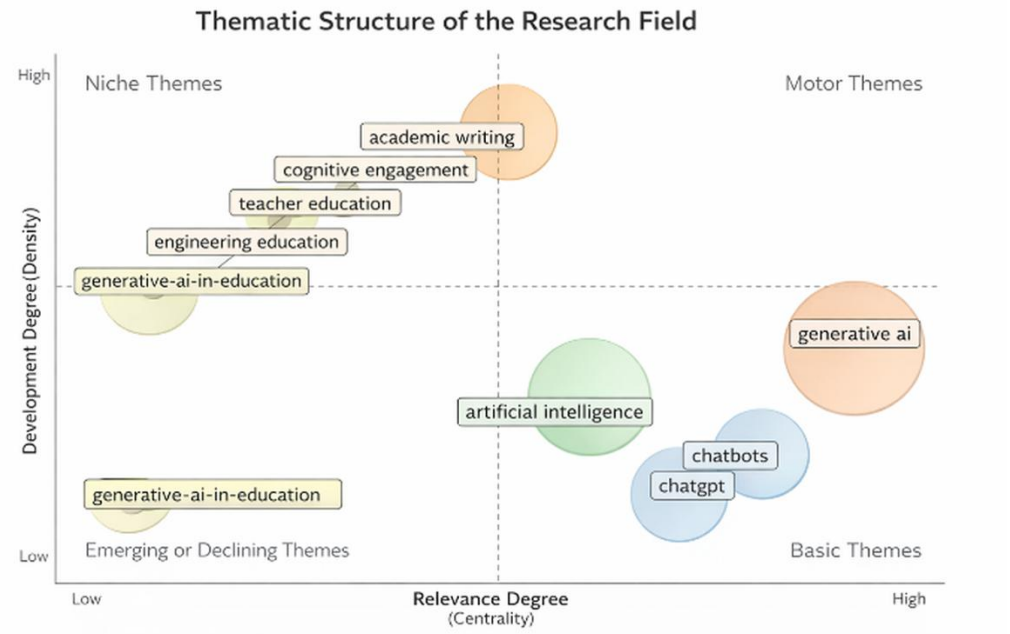


Figure 10. Strategic (thematic) map of generative artificial intelligence research in higher education. Produced in Biblioshiny from author keywords (250 terms; minimum cluster frequency of 5 per thousand documents; Louvain clustering). The horizontal axis represents centrality (relevance degree) and the vertical axis density (development degree), each computed as defined in Section 2.7 and split at the median of the observed cluster values. Bubble size is proportional to the number of keyword occurrences within the cluster. Cluster labels have been repositioned to avoid overlap; each theme occupies exactly one quadrant.

An examination of the thematic map presented in Figure 10 shows that the themes in the research field are positioned within four main categories: motor themes (high centrality–high density), basic themes (high centrality–low density), niche themes (low centrality–high density) and emerging/declining themes (low centrality–low density). This classification reveals both the significance of themes within the field and their level of conceptual maturity. The findings indicate that the theme of generative AI is positioned within the motor themes category. This theme exhibits both high centrality and high density, suggesting that it plays a decisive and

guiding role in the development of the research field. This demonstrates that generative artificial intelligence is both extensively studied and strongly connected with other themes in the literature. Within the basic themes category, concepts such as artificial intelligence, ChatGPT and chatbots are prominent. While these themes constitute the conceptual backbone of the field, they have not yet reached the same level of density as motor themes. This suggests that, although they are connected to a broad research network, they are still in the process of developing greater conceptual depth.

Among the niche themes, concepts such as academic writing, cognitive engagement, teacher education and engineering education stand out. Despite their high density, their low centrality indicates that they represent more specialized topics within specific subfields. This suggests that these themes have been studied in depth but maintain relatively limited interaction with the broader research network. In the category of emerging or declining themes, the theme of generative AI in education is identified. With low centrality and density, this theme represents a research area that is either in an early stage of development or addressed by a limited number of studies. However, considering the overall development trend of the field, this theme has the potential to evolve into a more central position in the future. One point of presentation requires clarification. In the strategic map, 'generative-ai-in-education' constitutes a single cluster with a single pair of centrality and density coordinates. Where its label appears more than once in the plotted output, this is a label-placement artefact of the plotting routine, which draws cluster labels with leader lines and may repeat a label, or position it outside the quadrant containing its bubble, when labels would otherwise overlap. The theme is nevertheless classified in a single quadrant emerging or declining on the basis of its centrality and density values relative to the medians defined in Section 2.7. Substantively, the coordinates of this theme place it close to both medians, which is itself informative: 'generative AI in education' is a theme in transition rather than a settled one. Given that two-thirds of the corpus was published in the final year of the window, the low-centrality, low-density quadrant should be read here as emerging rather than declining. Overall, the thematic structure analysis indicates that the research field is undergoing a dynamic and multilayered development process. Themes centered on generative artificial intelligence play a dominant and guiding role in the literature, while themes related to education, pedagogical applications and ethical dimensions are represented at varying levels of development.

3.10. Evolution of Research Themes

To reveal the transformation of themes and their developmental dynamics over time in the research field, a thematic evolution analysis was conducted. The results are presented in Figure 11.

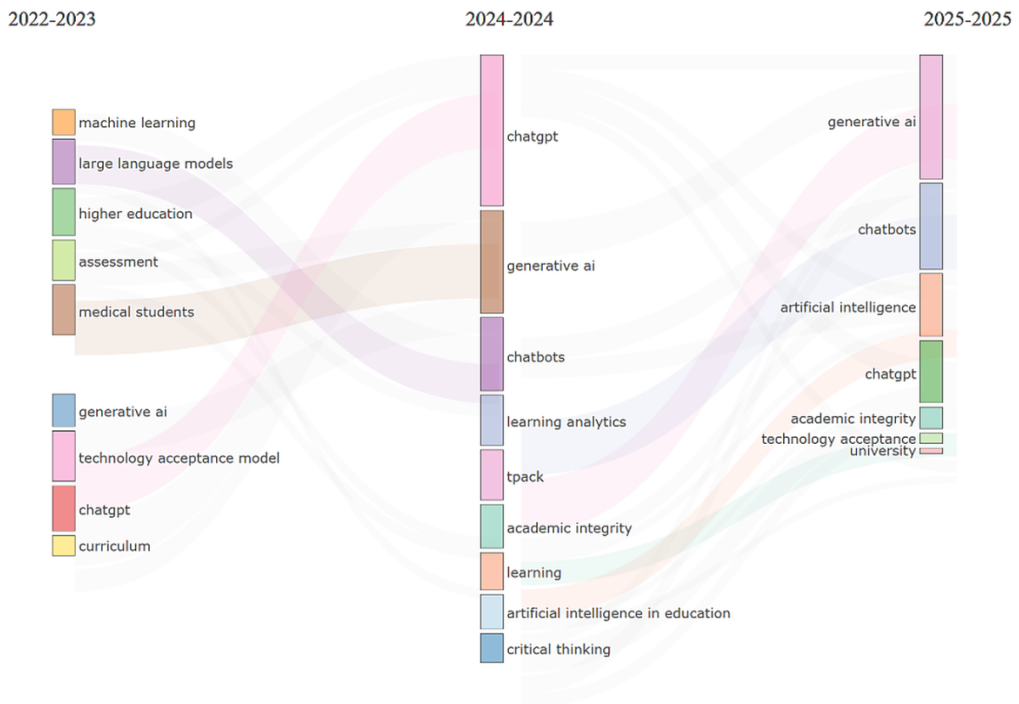


Figure 11. Thematic evolution of generative artificial intelligence research in higher education. Produced in Biblioshiny from author keywords using the parameters reported in Table 2, with cutting points at 2023 and 2024, yielding three periods (2022–2023, 2024 and 2025). Each column lists the themes identified within that period, labelled by the most frequent term in the cluster; bands link themes across adjacent periods by the inclusion index weighted by word occurrences and band width is proportional to that index. Block height is proportional to the weight of the theme within its period.

An examination of Figure 11 shows that the thematic structure of the field has changed substantially across the three periods and that the change follows a definite sequence rather than a simple accumulation of topics.

In the first period (2022–2023) the structure is fragmented and no theme dominates. Nine themes of broadly comparable weight are present, spanning three distinct concerns: a technical vocabulary (machine learning, large language models), the educational settings in which the technology was being examined (higher education, curriculum, medical students) and early evaluative and adoption concerns (assessment, technology acceptance model). ChatGPT and generative AI are already present but are no larger than the themes around them. This is the profile of a field still establishing what its object of study is.

In 2024 the structure consolidates sharply. ChatGPT becomes the largest theme by a clear margin, with generative AI second and the dispersion of the first period gives way to concentration around these two. The same period introduces themes that were absent earlier: academic integrity, learning analytics, TPACK, critical thinking and artificial intelligence in education. The combination is characteristic of a consolidation phase, in which a dominant object of study emerges and specialised research programmes begin to form around it.

In 2025 the dominant theme changes. Generative AI overtakes ChatGPT to become the largest theme, ChatGPT itself contracts and both chatbots and artificial intelligence grow. Academic integrity persists from the previous period, while technology acceptance and university appear as separate themes. The single most substantial flow in the whole diagram runs from ChatGPT in 2024 to generative AI in 2025.

That flow is the analytically significant feature of Figure 11 and it repays careful reading. It does not represent the replacement of one subject by another, since the two themes share much of their underlying vocabulary; it represents a change in how the field names its object. Researchers who described what they studied as ChatGPT in 2024 increasingly describe it as generative AI in 2025 a shift from a product to a category. This is a maturation signal rather than a terminological accident. A field that names its object after the tool it happens to use is tied to the fortunes of that tool, whereas one that names it by category can accumulate findings across successive tools. The persistence of academic integrity across the final two periods points the same way: a normative concern that has stabilised rather than passed. Overall, the thematic evolution analysis describes a field that moved from fragmentation in 2022–2023, through consolidation around a single product in 2024, to consolidation around a category in 2025, while developing a durable normative concern with academic integrity alongside its technical and pedagogical themes.

4. Discussion

The bibliometric analysis conducted in this study provides a comprehensive perspective on the scientific production trends, intellectual structure and thematic development of generative artificial intelligence research in higher education. This section is organised thematically and each subsection follows the same three-step structure: the finding obtained in this study is stated first, its analytical interpretation is developed second and the result is then compared with the existing literature in order to identify points of agreement, extension and divergence.

4.1. Growth Dynamics and the Age Structure of the Field

The production analysis established that output rose from 5 documents in 2022 to 2,001 in 2025, a 400-fold expansion in which 66.9% of the corpus appeared in the final year, while the annual growth rate fell successively from 3,900% to 293.0% and then to 154.6%. The combination of these two patterns continuing steep expansion in volume alongside a clear deceleration in rate is the analytically important result and it is one that raw publication counts obscure.

Interpreted structurally, this pattern indicates that the field has passed through its explosive entry phase and is now consolidating onto a steep but decelerating growth path. It also fixes the age structure of the literature: with 93.1% of documents published in the final two years, the great majority of this research has not yet been cited, replicated or contested. The practical consequence is that the present analysis maps an active research front rather than a settled knowledge base and that any claim about what the field has established, as distinct from what it is currently investigating, must be made with corresponding caution.

Read against the literature, the acceleration itself is consistent with earlier accounts. Kasneci et al. (2023) and Tlili et al. (2023) documented the abrupt redirection of educational technology research following the release of ChatGPT and Akpan et al. (2025) reported that 96% of the scientific output on conversational and generative artificial intelligence appeared between 2019 and 2023. The present findings extend these accounts in two respects. They demonstrate that the acceleration continued well beyond the windows those studies could observe and, more importantly, they show that the growth rate has begun to fall a transition that earlier maps, terminating in 2023 or 2024, were not positioned to detect.

4.2. Geographical Distribution and the Collaboration Deficit

The geographical analysis showed production concentrated in China (n = 890), the United States (n = 617) and Australia (n = 401), with substantial contributions from Indonesia, Malaysia, Saudi Arabia, Jordan and South Africa. Alongside this, however, the corpus records an international co-authorship rate of 10.11% against a collaboration index of 2.83 and only

17.1% single-authored documents.

The interpretation of these figures turns on reading them jointly. Co-authorship is clearly the norm in this field, but it is overwhelmingly domestic: roughly nine in ten collaborative papers are produced within a single national system. This constitutes a substantive limitation on what the field can currently establish. The questions that dominate the literature assessment security, academic integrity policy, equity of access, disciplinary norms of permissible use are institutionally and nationally contingent and answering them requires precisely the cross-system comparison that current collaboration patterns do not support. The prominence of China, the United States and Australia is plausibly attributable to research infrastructure, technological investment and digital transformation policy in higher education (Coulter-Smith, 2022; Kavitha et al., 2024; Nugroho, 2024), while the visible contribution of Indonesia, Malaysia, Saudi Arabia, Jordan and South Africa indicates that the field is not the preserve of a small group of developed systems (Dai et al., 2026).

Compared with the literature, this finding both confirms and qualifies earlier work. Zawacki-Richter et al. (2019) observed that artificial intelligence research in higher education concentrates in countries with strong research infrastructures and Metli (2023) and Ng and Ho (2025) report similarly uneven distributions. Akpan et al. (2025), however, characterised the intellectual structure of conversational and generative artificial intelligence research as exhibiting strong collaboration. The present analysis suggests that this characterisation requires refinement when the field is narrowed to higher education: collaboration is indeed strong, but it is national rather than international and the distinction is consequential for the kinds of knowledge the field is able to produce.

4.3. The Position of Türkiye in the Field

The country-level analysis placed Türkiye in the corpus with 96 publications, in the same output band as the United Arab Emirates, Peru, Hong Kong and South Africa, but below Saudi Arabia (n = 208), Indonesia (n = 186), Malaysia (n = 164), Taiwan (n = 144), India (n = 139), Jordan (n = 135) and Germany (n = 153).

Two interpretations of this position must be distinguished. The first is substantive: measured against regional peers with higher education systems of comparable or smaller scale, Turkish output in this specific area is lower than the size of the system would lead one to expect and the shortfall lies with those peers rather than with the global leaders. The second is methodological and works in the opposite direction. A substantial part of Turkish scholarship on generative artificial intelligence in higher education is published in Turkish, in nationally indexed journals that fall outside the Web of Science and Scopus coverage applied here and outside the English-language restriction. The figure of 96 is therefore a lower bound on Turkish activity in this area rather than a complete measure of it and the same qualification applies to every non-Anglophone system in the dataset.

The national literature supports this reading. Metli (2023) and Tonbuloğlu and Avcı Akbel (2023) demonstrate an established Turkish bibliometric tradition at the intersection of educational technology and artificial intelligence and Büyükeke (2025) reports empirical evidence from 305 Turkish undergraduates showing that perceived usefulness and perceived ease of use predict intention to use generative artificial intelligence tools in programming courses, with perceived trust operating on both antecedents. The theoretical significance of that finding for the present analysis is that technology acceptance appears in the conceptual structure map (Figure 9) as a component of the pedagogical-transformation cluster and the Turkish evidence indicates that the mechanism operates in a non-Anglophone system in substantially the same way. This is precisely the kind of cross-system corroboration that the field's low international co-authorship rate currently leaves underexploited.

4.4. Citation Structure and the First-Mover Effect

The citation analysis established that the ten most-cited documents account for 8.75% of all citations generated by the corpus and the three most-cited for 5.4%, from 0.33% and 0.10% of documents respectively.

This concentration is best interpreted as a first-mover effect rather than as a settled quality signal. All three leading documents were published in 2023 or earlier and address academic integrity or the broad pedagogical implications of ChatGPT; in a corpus where two-thirds of the literature appeared in 2025, early publication confers a decisive and largely mechanical citation advantage. Moreover, the works that appeared first were framing and agenda-setting pieces rather than empirical investigations, which means the field's citation structure currently rewards early problem definition rather than accumulated evidence.

This reading is consistent with the content of the most-cited works themselves. Kasneci et al. (2023) set out the opportunities and challenges of large language models for education as a programmatic agenda and Cotton et al. (2024) argued that generative artificial intelligence in academic writing generates both learning opportunities and integrity risks. Both are framing contributions and both were early. The implication for readers of bibliometric maps of this field is that the highly cited core should be read as evidence of what the field first agreed to worry about, not as evidence of where its most robust findings now lie.

4.5. Intellectual and Conceptual Structure

The co-occurrence network built on indexed vocabulary returned four clusters organised by disciplinary setting instructional technology, health-professions education, technology acceptance and a general artificial intelligence core while correspondence analysis on harmonised author keywords returned four clusters organised by conceptual concern: technological

development, educational application, ethics and assessment and pedagogical transformation.

The substantive interpretation is that these two solutions describe different cross-sections of the same field rather than competing accounts of it. Read together, they show that the literature is stratified along two axes at once: by the disciplinary setting in which the research is conducted and by the conceptual concern that motivates it. The most consequential finding to emerge from the comparison is the prominence of health-professions education in the indexed solution, which suggests that a substantial and coherent body of work on generative artificial intelligence in medical and nursing education has developed with its own vocabulary and its own indexing footprint and which existing bibliometric reviews of this field have not reported. That this subfield is invisible in the author-keyword solution is itself instructive: it is identifiable through how the literature is classified rather than through how its authors describe it.

The three-axis architecture is itself consistent with the wider literature. Zawacki-Richter et al. (2019) identified a comparable structure at the intersection of artificial intelligence and education and Dúo-Terrón (2024), Ng and Ho (2025) and Guo et al. (2026) report similar technological, applied and ethical groupings. The contribution of the present analysis is not the discovery of this architecture but its verification: because the parameters of both analyses are reported in Table 2, the structure can be reproduced, tested at other thresholds and compared against future corpora.

4.6. Thematic Structure and Thematic Evolution

The strategic map located generative AI as the sole motor theme; artificial intelligence, ChatGPT and chatbots as basic themes; academic writing, cognitive engagement, teacher education and engineering education as niche themes; and generative AI in education in the emerging or declining quadrant. The evolution analysis traced a movement from a fragmented set of technical, contextual and adoption themes in 2022–2023, through consolidation around ChatGPT in 2024, to the ascendancy of generative AI as the dominant theme in 2025, with academic integrity establishing itself as a durable concern across the final two periods.

Interpreted together, these results describe a field whose centre of gravity has shifted from asking what these technologies are to asking what they do to educational practice and to its normative order. The appearance of academic writing, cognitive engagement and teacher education as high-density but low-centrality themes indicates that specialised research communities have formed around them, but that these communities remain weakly connected to the core of the field. The presence of only one theme in the motor quadrant is equally notable: it suggests that the field's integrative capacity is currently concentrated in a single concept.

Against the literature, this trajectory is well corroborated. Bozkurt (2023) and Tlili et al. (2023) documented the early movement of educational research from technological description towards pedagogical and ethical analysis; Plata et al. (2023) and Bittle and El-Gayar (2025) traced the emergence of academic integrity as a distinct research and policy theme; and Ng and Ho (2025) reported comparable thematic diversification. The present analysis adds temporal precision to these accounts by locating the transition within specific time slices and by reporting the parameters under which the thematic partition was obtained.

4.7. A Critical Reading: Three Structural Risks

Taken together, the findings support a more critical reading of this field than descriptive maps typically offer and three structural risks warrant explicit statement.

The first is ChatGPT-centrism. ChatGPT occupies a central position in every analysis reported here, appearing as a basic theme, as a dominant node in the co-occurrence network and as a defining term in the conceptual structure. A field whose conceptual architecture is organised around a single commercial product is exposed to a specific form of obsolescence: findings framed in terms of a particular tool do not straightforwardly transfer when that tool is superseded. The corresponding research implication is that studies should be designed around transferable constructs the pedagogical function performed, the cognitive process affected, the integrity risk incurred rather than around the current market leader.

The second is the collaboration deficit documented in Section 4.2. An international co-authorship rate of 10.11% in a field whose central questions are institutionally contingent implies a real risk of parallel national literatures converging on similar findings independently, without the comparative design that would make those findings cumulative.

The third is early thematic saturation. The evidence for this is structural rather than impressionistic: a single theme occupies the motor quadrant; the basic quadrant contains the field's most frequent and most general terms; and the niche quadrant holds several well-developed but poorly connected themes. This configuration a broad, shallow core surrounded by disconnected specialised pockets combined with a strongly right-skewed author distribution in which no dominant research school has formed, is consistent with a field producing a high volume of thematically similar work rather than a deepening cumulative programme. The very flat author productivity distribution reported in Section 3.3 reinforces this reading, since a literature built largely from single contributions is less likely to develop sustained lines of inquiry.

These three risks are not arguments against research on generative artificial intelligence in higher education. They are arguments about how the next phase of it should be designed: around transferable constructs rather than named tools, across national systems rather than within them and towards the consolidation of existing themes rather than the continued proliferation of new ones.

5. Limitations

The findings of this study should be read in the light of several limitations, which fall into three groups: limitations of corpus coverage, limitations of the analytical procedure and limitations inherent to bibliometric inference.

Coverage. The corpus was constructed exclusively from the Web of Science Core Collection and Scopus. Although these are the two most widely used citation databases and their combination provides broader coverage than either alone, other sources were not searched, including ERIC, Dimensions, Lens, Google Scholar, regional indexes and national indexes such as TR Dizin. Education-specific and regionally indexed outlets are therefore systematically under-represented. Second, only English-language publications were included, which excludes a substantial body of non-Anglophone scholarship; as noted in Section 4.3, country-level counts for non-Anglophone systems, including the figure of 96 reported for Türkiye, should accordingly be treated as lower bounds rather than complete measures. Third, only articles and reviews were included. This excludes conference proceedings, which are a primary publication venue in computer science and therefore carry a meaningful share of early technical work on large language models, as well as book chapters, editorials and preprint servers such as arXiv. Fourth, although the rationale for the search vocabulary is set out in Section 2.3, records that name only a specific model Gemini, Claude, Llama, Copilot or DeepSeek without employing any umbrella term would not have been retrieved; a model-name-inclusive sensitivity search is recommended for future replications. Fifth, the search was executed on 10 March 2026 and the window closes at 2025, so publications from 2025 indexed after that date are absent and the 2025 count is a lower bound.

Procedure. As set out in Section 2.4, eligibility was enforced through the search query and the database filters rather than through a manual title-and-abstract screening stage. This maximises reproducibility, but it means that precision was not improved by human judgement: records in which a university appears only as an author affiliation and records in which generative artificial intelligence is mentioned only incidentally, remain in the corpus and will have contributed to the keyword and citation distributions reported here. As a single-authored study, all design decisions were made by one researcher and no inter-rater reliability coefficient is applicable. Author disambiguation, institutional attribution and keyword indexing rely on the metadata supplied by the source databases, so homonymous author names, name variants and inconsistent affiliation strings may affect the productivity rankings. Country production was computed by full counting on author affiliations, meaning that a document authored across several countries contributes one unit to each; country counts therefore exceed the document total and must not be summed. A further constraint applies specifically to Figure 8. Because that network deliberately retains database index keywords, demographic and study-design descriptors such as human, female, adult, article and questionnaire appear as network nodes although they are indexing conventions rather than research themes and variant forms of the same term artificial intelligence and artificial-intelligence, student and students, human and humans appear as separate nodes. This is inherent to index-term analysis and constrains the resolution obtainable from that figure; it does not affect the author-keyword analyses reported in Figures 9 to 11, to which the harmonisation protocol of Section 2.5 was applied in full. Finally, the conceptual and thematic solutions are conditional on the parameters reported in Table 2 the data field, the occurrence thresholds, the normalisation and the clustering algorithm and alternative settings may yield partitions that differ in number or composition. These parameters are reported in full precisely so that such sensitivity can be tested by others.

Inference. Bibliometric analysis describes the publication, citation and vocabulary behaviour of a literature; it does not assess the methodological quality, validity or practical effect of the studies it maps. High citation counts index visibility rather than rigour and in a corpus of this age structure they index publication date to a considerable degree, as Section 4.4 demonstrates. Similarly, thematic centrality and density measure the position of a term within a keyword network, not the importance of the underlying educational phenomenon. The map presented here should therefore be understood as a description of how this research field is currently organised and not as an evaluation of what it has established.

6. Conclusion

This study mapped the scientific production structure, intellectual and conceptual organisation and thematic evolution of generative artificial intelligence research in higher education, using a dual-database corpus of 2,992 publications and reporting the full analytical parameters of every procedure applied.

Four conclusions follow. First, the field is expanding steeply but at a decelerating rate and its age structure with 93.1% of documents published in the final two years means that it currently constitutes an active research front rather than a consolidated knowledge base. Second, the field is stratified along two axes simultaneously: by disciplinary setting, where the indexed vocabulary separates an instructional-technology tradition from a substantial and previously unreported health-professions education tradition and by conceptual concern, where author keywords separate technological, educational-application, ethical-assessment and pedagogical dimensions. Neither view is reducible to the other and recovering both is what triangulating two independent clustering procedures makes possible. Third, generative AI is the field's single motor theme and its thematic trajectory runs from technological

exploration towards pedagogical integration and academic integrity. Fourth, the field exhibits three structural risks ChatGPT-centrism, an international co-authorship rate of only 10.11% and early indications of thematic saturation that its rapid growth has tended to conceal.

The theoretical contribution of the study lies in demonstrating that generative artificial intelligence research in higher education is not a technological literature with pedagogical applications appended, but an interdisciplinary domain in which technological, pedagogical and normative concerns are structurally co-constitutive; both clustering procedures recover the normative dimension as an independent axis rather than as a subsidiary theme. Its methodological contribution is a reproducible protocol: the parameters in Table 2 allow every result reported here to be re-derived, tested at alternative thresholds, or extended to subsequent time windows.

For higher education institutions, the findings indicate that policy on generative artificial intelligence cannot be confined to detection and prohibition, since the literature has already moved towards integration, assessment redesign and instructor development. Institutional strategy should accordingly address academic integrity policy, assessment security and professional development jointly rather than sequentially.

For future research, three directions follow directly from the structural risks identified. Studies should be designed around transferable pedagogical and cognitive constructs rather than around named commercial tools; cross-national comparative designs are needed to convert parallel national literatures into cumulative evidence; and longitudinal work is required to establish the long-term effects of these tools on critical thinking, creativity and academic productivity, which the field's current age structure makes it impossible to assess. Replications extending the corpus to conference proceedings, to non-English publications and to model-name-inclusive search vocabularies would additionally test the boundaries of the map presented here.

Author Contributions

The percentages of the author' contributions are presented below. The author reviewed and approved the final version of the manuscript.

	Ş.B.
C	100
D	100
S	100
DCP	100
DAI	100
L	100
W	100
CR	100
SR	100
PM	100
FA	100

C= concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

Conflict of Interest

The author declared that there is no conflict of interest.

Ethical Consideration

This study is grounded in a bibliometric examination of previously published scholarly works. It does not involve the collection of data from human participants, nor does it utilize personal data or pose any ethical concerns. Accordingly, obtaining ethical approval was deemed unnecessary for the conduct of this research.

Acknowledgements

The author would like to thank the developers of the Web of Science Core Collection and Scopus databases for providing access to high-quality bibliographic data used in this study. No additional acknowledgments are declared.

Data Availability Statement

The bibliographic records analysed in this study were retrieved from the Web of Science Core Collection and Scopus, which are subscription-based databases; the underlying records are therefore available to any researcher with institutional access, using the search queries, filters and retrieval date reported in Section 2.3. The merged and cleaned dataset, the VOSviewer thesaurus file and the analysis scripts are available from the corresponding author on reasonable request.

References

Akpan, I. J., Kobara, Y. M., Owolabi, J., Akpan, A. A., & Offodile, O. F. (2025). Conversational and generative artificial intelligence and human–chatbot interaction in education and research. *International Transactions in Operational Research, 32*(3), 1251–1281. https://doi.org/10.1111/itor.13522

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

Baek, C., & Doleck, T. (2020). A bibliometric analysis of the papers published in the *Journal of Artificial Intelligence in Education* from 2015–2019. *International Journal of Learning Analytics and Artificial Intelligence for Education*, 2(1), 67–84. https://doi.org/10.3991/ijai.v2i1.14481

Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, 16(4), 296. https://doi.org/10.3390/info16040296

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment, 2008(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bozkurt, A. (2023). Unleashing the potential of generative AI, conversational agents and chatbots in educational praxis: A systematic review and bibliometric analysis of GenAI in education. *Open Praxis, 15*(4), 261–270. https://doi.org/10.55982/openpraxis.15.4.609

Büyükeke, A. (2025). Öğrencilerin programlama derslerinde üretken yapay zekâ araçlarını kabulü: Genişletilmiş teknoloji kabul modeli ile Türkiye'den bulgular [Students' acceptance of generative artificial intelligence tools in programming courses: Findings from Türkiye with an extended technology acceptance model]. *Journal of University Research, 8*(4), 543–556. https://doi.org/10.32329/uad.1728785

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics, 22*(1), 155–205. https://doi.org/10.1007/BF02019280

Chiu, T. K. F. (2024). Future research recommendations for transforming higher education with generative AI. *Computers and Education: Artificial Intelligence*, 6, 100197. https://doi.org/10.1016/j.caeai.2023.100197

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics, 5*(1), 146–166. https://doi.org/10.1016/j.joi.2010.10.002

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

Coulter-Smith, L. (2022). The AI economy and higher education. In A. Kaplan (Ed.), Digital transformation and disruption of higher education (Chapter 29). Cambridge University Press. https://doi.org/10.1017/9781108979146.037

Dai, K., Liu, Y., & Zhang, X. (2026). Generative AI in higher education: A bibliometric review of emerging trends, power dynamics and global research landscapes. *Computers and Education: Artificial Intelligence, 10, 100544.* https://doi.org/10.1016/j.caeai.2026.100544

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

Dúo-Terrón, P. (2024). Generative artificial intelligence: Educational reflections from an analysis of scientific production. *Journal of Technology and Science Education, 14*(3), 756–769. https://doi.org/10.3926/jotse.2680

Flores-Velásquez, C. H., Olivares-Zegarra, S., Dávila-Ignacio, C., Arévalo-Tuesta, J. A., Morales-Romero, G., Trinidad-Loli, N., Caycho-Salas, B., Aybar-Bellido, I., Arones, M., & Aldana-Trejo, F. (2024). A bibliometric review of studies about the acceptance of artificial intelligence technologies in teaching and learning in higher education. *International Journal of Learning, Teaching and Educational Research, 23*(3), 275–292. https://doi.org/10.26803/ijlter.23.3.14

Francis, N. J., Jones, S., & Smith, D. P. (2025). Generative AI in higher education: Balancing innovation and integrity. *British Journal of Biomedical Science*, 81, 14048. https://doi.org/10.3389/bjbs.2024.14048

Guo, J., Kang, E. K. M. S., & Ghazali, N. (2026). Generative artificial intelligence in education from 2021 to 2025: A scientometric review. *Journal of Curriculum and Teaching, 15*(1), 102–118. https://doi.org/10.5430/jct.v15n1p102

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … 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

Kavitha, K., Joshith, V. P., Rajeev, N. P., & Asha, S. (2024). Artificial intelligence in higher education: A bibliometric approach. *European Journal of Educational Research, 13*(3), 1121–1137. https://doi.org/10.12973/eu-jer.13.3.1121

Lodge, J. M., Thompson, K., & Corrin, L. (2023). Mapping out a research agenda for generative artificial intelligence in tertiary education. *Australasian Journal of Educational Technology*, 39(1), 1–8. https://doi.org/10.14742/ajet.8695

Metli, A. (2023). Articles on education and artificial intelligence: A bibliometric analysis. Journal of Social Sciences and Education, 6(Education Special Issue), 279–312. https://doi.org/10.53047/josse.1352197

Muskhir, M., Luthfi, A., Watrianthos, R., Usmeldi, U., Fortuna, A., & Samala, A. D. (2024). Emerging research on virtual reality applications in vocational education: A bibliometric analysis. *Journal of Information Technology Education: Innovations in Practice*, 23, 1–26. https://doi.org/10.28945/5284

Ng, S.-L., & Ho, C.-C. (2025). Generative AI in education: Mapping the research landscape through bibliometric analysis. *Information, 16*(8), 657. https://doi.org/10.3390/info16080657

Nugroho, L. A. (2024). Exploring research trends of ChatGPT in education through bibliometric analysis. *Scholaria: Jurnal Pendidikan dan Kebudayaan, 14*(2), 98–106. https://doi.org/10.24246/j.js.2024.v14.i2.p98-106

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*, n71. https://doi.org/10.1136/bmj.n71

Plata, S., De Guzman, M. A., & Quesada, A. (2023). Emerging research and policy themes on academic integrity in the age of ChatGPT and generative AI. *Asian Journal of University Education*, 19(4), 743–758. https://doi.org/10.24191/ajue.v19i4.24697

Prayuda, M. S., Ginting, F. Y. A., & Tamba, L. (2026). Tracing two decades of artificial intelligence in education: A bibliometric analysis of trends, themes and future directions (2000–

2025). *European Journal of Educational Research*, 15(1), 285–304. https://doi.org/10.12973/eu-jer.15.1.285

Samala, A. D., Rawas, S., Wang, T., Reed, J. M., Kim, J., Howard, N.-J., & Ertz, M. (2025). Unveiling the landscape of generative artificial intelligence in education: A comprehensive taxonomy of applications, challenges and future prospects. *Education and Information Technologies, 30(3), 3239–3278.* https://doi.org/10.1007/s10639-024-12936-0

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(1), 15. https://doi.org/10.1186/s40561-023-00237-x

Tonbuloğlu, B., & Avcı Akbel, B. (2023). Research trends in the field of emergency remote teaching: A bibliometric analysis of open access literature. *Online Learning*, 27(1), 485–517. https://doi.org/10.24059/olj.v27i1.3241

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/s11192-009-0146-2

Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics, 4*(4), 629–635. https://doi.org/10.1016/j.joi.2010.07.002

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

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*(3), 429–472. https://doi.org/10.1177/1094428114562629