PILLAR OF PHYSICS EDUCATION
Vol. 19 (4), 2026, page.464-478
ISSN (Print) : 2337-9618
ISSN (Online) : 2684-8880

SCIENTIFIC MAPPING OF THE IMPLEMENTATION OF GENERATIVE ARTIFICIAL INTELLIGENCE TECHNOLOGY IN SCIENCE LEARNING: A BIBLIOMETRIC ANALYSIS

Irrefyna Alit Seyruni[1], Asrizal[1*], Harman Amir[1], Dea Stivani Suherman[1]

[1] *Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Padang, Jl. Prof. Dr. Hamka Air Tawar Padang 25131, Indonesia*
Corresponding author. Email: asrizal@fmipa.unp.ac.id

ABSTRACT

This study maps research on the use of Generative Artificial Intelligence (GenAI) in science learning, operationally defined here as teaching and learning documents in the subject areas of Computer Science, Physics and Astronomy, and Earth and Planetary Sciences. A bibliometric analysis was conducted on 1,400 Scopus-indexed articles and conference papers published in 2023–2025, retrieved with the keywords "generative ai" and "teaching", using performance analysis (Scopus Analyze Results) and keyword co-occurrence mapping (VOSviewer). Publications grew from 54 in 2023 to 998 in 2025. The keywords formed four thematic clusters: generative AI and higher education; teaching and computing education; intelligent learning systems and computer-aided instruction; and artificial intelligence, teachers, and AI literacy. The overlay suggests, only tentatively, that more recent topics lean toward computer-aided instruction and computing education, whereas ethics, AI literacy, and teacher education appear less prominent. The United States and China are the largest contributors. The main limitation is that the corpus is almost entirely Computer Science (1,380 of 1,400 documents), so "science learning" here largely means GenAI in computing-related teaching, and the findings cannot be generalized to other science domains. Possible gaps, to be tested with broader searches, include physics-specific topics, teacher readiness, ethical use, and the contexts of developing countries, including Indonesia.

Keywords: Bibliometric analysis; Generative AI; Science learning; Scopus; VOSviewer

Pillar of Physics Education is licensed under a Creative Commons Attribution ShareAlike 4.0 International License.

I. INTRODUCTION

Digital technology has changed how teachers and students interact with learning resources, shifting from information delivery toward support for personalization, feedback, assessment, and lesson planning. One such technology is *Artificial Intelligence* (AI), which can serve as a subject, a mediator, or a supporting tool for teacher–student interaction [1]. This development has been accelerated by *Generative Artificial Intelligence* (GenAI), a technology that produces new content based on user *prompts*. The emergence of ChatGPT in late 2022 widened the use of GenAI in education. Ideally, its use is accompanied by the ability to evaluate the accuracy of information, to understand the limitations of its outputs, and to use the technology responsibly [2]. Previous mapping studies show that research on GenAI in education covers technology applications, ethics, user perspectives, adoption, and technology performance [3].

GenAI is important in science learning, which demands conceptual understanding, problem solving, and higher-order thinking. The application of AI in science learning has been associated with improved conceptual understanding and scientific literacy [4], and digital technology in physics learning has likewise been shown to support students' conceptual understanding and new literacies [5]. A randomized controlled trial reported that an AI tutor designed on pedagogical principles produced better learning outcomes than in-class active learning [6], while ChatGPT-based feedback improved students' achievement in a physics laboratory course [7]. For

educators, GenAI helps in designing lesson units, rubrics, and quizzes [8] and has been tested in physics task development by prospective teachers [9].

The benefits of GenAI, namely personalized feedback, responsive tutoring, and task-design support, are not automatic. Unguarded access to GenAI can harm high school students' learning in mathematics [10]; all ChatGPT answers to mechanics comprehension questions were found to be incorrect, inaccurate, incomplete, or misleading [11]; and eleventh-grade senior high school students in Indonesia tended to accept GenAI answers quickly in physics learning [12]. GenAI is therefore best used with pedagogical design, critical evaluation of its outputs, and AI literacy. Such evidence shows that the value of GenAI depends on how it is designed into learning and on how its outputs are examined by students and teachers.

Attention to these opportunities and challenges is evident in systematic reviews. The application of AI in science learning has developed in various forms to support adaptive learning [13], and the implementation of AI in science and chemistry education spans learning, laboratories, personalization, teacher–student interaction, and assessment [14]. In the GenAI domain, a review of the relationship between GenAI and science education has been conducted [15], while a review of 85 studies (49 of which entered the meta-analysis) found a positive effect of GenAI on STEM cognitive learning outcomes that was highly heterogeneous and largely attributable to publication bias [16]. These reviews show that research attention has grown, but they synthesize substantive findings rather than map the structure and direction of the field.

Bibliometric approaches have also been used to map publication patterns and thematic structures. In science education, the relationships among AI research topics have been mapped with VOSviewer [17]; AI research in physics education over two decades has been mapped in the Scopus database [18]; and 3,217 AI studies from 2015–2024 have been analyzed [19]. For GenAI, publications and themes in education have been mapped [20], while other mapping studies focus on higher education [21,22]. Existing mappings thus cover AI in general, GenAI in education in general, or, in the closest case, a single tool (ChatGPT) in science education.

This study adopts five criteria, which are the authors' own design choices rather than an established standard, for a mapping of a rapidly developing field: (1) trends and directions of development, (2) topics that have been widely studied and those that remain rare, (3) the knowledge structure among keywords, (4) country contributions and subject-area distribution, and (5) a comprehensive and objective picture despite the continuously growing literature. However, the mappings reviewed here do not yet meet these criteria for GenAI in science learning. First, trends are difficult to identify because existing mappings cover AI in general in science education or GenAI in education in general and in higher education. Second, research gaps have not been comprehensively portrayed because previous studies address separate aspects, such as task and quiz design, challenges of AI in science education, and cognitive learning outcomes. Third, the knowledge structure of GenAI in science learning has not been mapped because earlier mappings cover AI in general. Fourth, information on country contributions and subject-area distribution remains limited, while studies of international collaboration focus on higher education. Fifth, the GenAI literature grows rapidly, so subjective narrative reviews are less able to accommodate its volume objectively [23]. The closest existing mappings are a Scopus-based visualization of ChatGPT research in science education for 2022–2024, which covers a single tool and a different period [24], and a Web of Science-based mapping of AI in physics education for 2021–2025, which covers AI in general [25]. To the best of the authors' knowledge, based on a Scopus search on 14 October 2026 with the query TITLE-ABS-KEY(("generative ai" OR chatgpt) AND (bibliometric OR "science mapping") AND ("science education" OR "physics education" OR "science learning")), no mapping covers GenAI in teaching and science learning for 2023–2025 in terms of keyword structure, subject areas, and country contributions.

Bibliometric analysis is a quantitative method for evaluating the development of a field based on publication data, so that trends, thematic structures, and research gaps can be identified systematically [26]. The method processes large volumes of bibliographic data in a measurable and verifiable manner. The number of publications per year describes growth trends, *keyword co-occurrence* analysis in VOSviewer maps the knowledge structure and topic development [27], and the *Analyze Results* feature in Scopus presents the subject-area distribution and country contributions. This approach complements systematic reviews, which focus on synthesizing substantive findings, and has been used to map GenAI publications and themes in education.

The aim of this study is to map the scientific development of the use of GenAI in science learning based on Scopus-indexed publications from 2023–2025. Science learning is defined operationally through subject-area filters, and because the resulting corpus is dominated by Computer Science, the mapping mainly describes GenAI in computing-related teaching and uses it as a basis for identifying gaps in science and physics learning. This focus distinguishes it from previous mappings that cover AI in general or GenAI in education in general. The study addresses seven research questions: (1) What is the growth pattern of publications? (2) How are

keywords interrelated in the network visualization? (3) How have topics developed in the overlay visualization? (4) Which topics dominate the density visualization? (5) How are subject areas distributed? (6) How do countries contribute? and (7) What research gaps emerge, particularly for physics-specific topics? The results are expected to help science and physics education researchers determine directions for future research.

II. METHOD

This study employed bibliometric analysis with a descriptive quantitative approach, which processes publication data numerically to describe and monitor the development of a field without testing hypotheses or causal relationships. Because it relies on traceable metadata, its procedures and results are transparent and easy to verify. The method identifies publication growth, keyword distribution, thematic structure, subject areas, and country contributions in research on GenAI in science learning. The analysis therefore describes the structure and direction of the literature and does not evaluate the effectiveness of GenAI in learning.

The data were secondary publication metadata from Scopus, chosen because of its broad journal coverage and its curated and verified data [28]. The period was limited to 2023–2025, the first three years after ChatGPT appeared in late 2022. Only complete calendar years were included: 2026 was excluded because it was still incomplete when the data were downloaded, so its partial count would not be comparable with the full-year counts. The data were downloaded on 14 October 2026.

The search used the *TITLE-ABS-KEY* field with the keywords "*generative ai*" AND "*teaching*" so that only documents containing both concepts were retrieved. The keyword "*teaching*" represents teaching and learning activities, whereas the scope of science was approached by restricting subject areas at the screening stage. The operational definition of "science learning" used in this study is set out in the next paragraph. Computer Science was selected because it is the main domain of GenAI development and evaluation, whereas Physics and Astronomy and Earth and Planetary Sciences represent the natural science disciplines as learning contexts. Because Computer Science accounts for 1,380 of the 1,400 documents (about 98.6%), the corpus is essentially GenAI in computing-related teaching; the findings are interpreted as such and are used to draw implications for science and physics learning rather than to describe science learning as a whole. The filter also excludes documents classified only under Social Sciences, where most education and science-education journals are indexed; this is a limitation of the design and is discussed in the Limitations. The results were screened sequentially by publication year (*PUBYEAR* 2023–2025), subject area (*SUBJAREA*: Computer Science, Physics and Astronomy, and Earth and Planetary Sciences), and document type (*DOCTYPE*: *article* and *conference paper*); reviews, book chapters, and other document types were excluded so that the corpus consists of primary research outputs. The keyword search returned 5,150 documents; 2,832 remained after the period filter, 1,765 after the subject-area filter, and 1,400 after the document-type filter. These 1,400 documents remained and served as the main data, with metadata exported in a VOSviewer-compatible format. The complete Scopus query string was: TITLE-ABS-KEY(" generative ai " AND " teaching ") AND (LIMIT-TO (SUBJAREA,"COMP") OR LIMIT-TO (SUBJAREA,"EART") OR LIMIT-TO (SUBJAREA,"PHYS")) AND (LIMIT-TO (DOCTYPE,"cp") OR LIMIT-TO (DOCTYPE,"ar")) AND PUBYEAR > 2022 AND PUBYEAR < 2026. No language filter was applied. The exported Scopus metadata and the thesaurus file are described in the Data Availability statement.

Operational definition of science learning. In this study, a document is treated as belonging to GenAI in science learning only when it meets all four of the following criteria: (1) its title, abstract, or keywords contain both "generative ai" and "teaching", where "generative ai" identifies the technology and "teaching" stands for teaching and learning activities; (2) Scopus classifies it under at least one of three subject areas, namely Computer Science (COMP), Physics and Astronomy (PHYS), or Earth and Planetary Sciences (EART), which represent, respectively, the field in which GenAI is developed and evaluated and the natural science disciplines that serve as learning contexts; (3) it is an article or a conference paper; and (4) it was published between 2023 and 2025. No science-specific terms, such as "physics", "chemistry", "biology", "science education", or "STEM", were used in the query, so a document is not required to discuss science content to be included, and science enters the definition only through the subject-area classification assigned by Scopus. "Science learning" is therefore a delimited, operational construct and not a comprehensive definition of the field; the consequences of this choice for interpretation are discussed in the Discussion and Limitations.

The procedure adapted the bibliometric workflow of Zupic and Čater [29] into five stages: data collection, data processing, bibliometric analysis, data visualization, and interpretation (Figure 1). Metadata were retrieved from Scopus according to the search strategy and then processed with Microsoft Excel and Notepad. A total of

7,023 keywords were extracted from all documents that passed screening, then normalized using a thesaurus file to unify singular–plural forms, abbreviations, and spelling variants, leaving 6,882 keywords ready for analysis. The thesaurus covered only these three types of variants, which explains the small reduction (141 keywords, about 2%). Labels such as "education computing", "personnel training", and "high educations" resemble Scopus index keywords, which reflect Scopus's classification vocabulary rather than the authors' own wording. The analysis used all keywords (author and index keywords) from the exported records, which is why Table 1 contains both author-style keywords (e.g., "ai literacy" and "chatgpt") and Scopus index keywords.

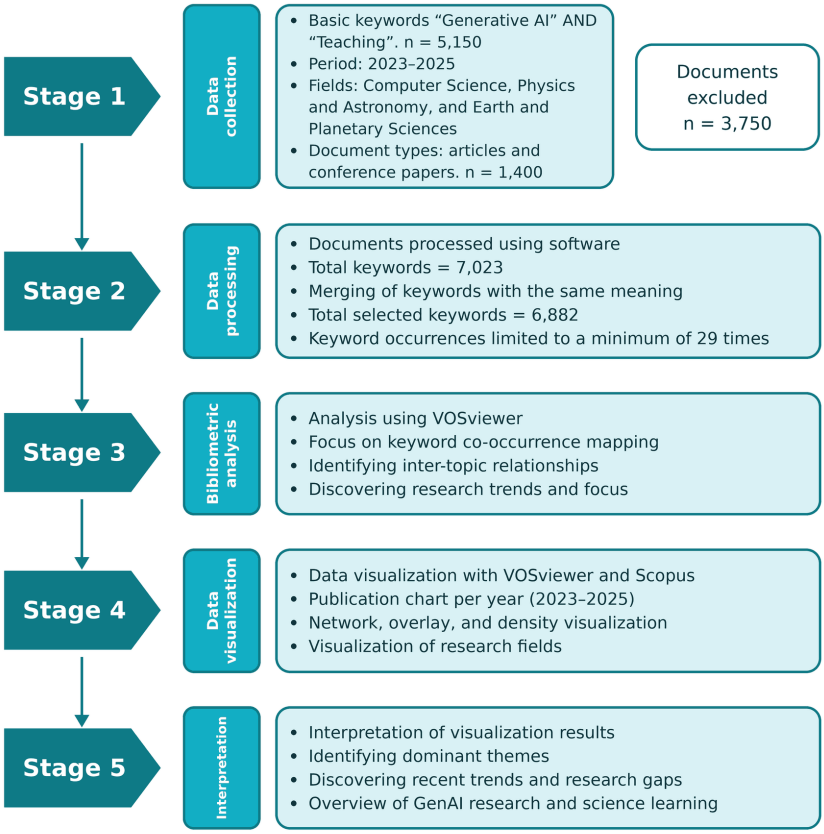


Figure 1. Stages of the bibliometric study

Science mapping used *keyword co-occurrence* analysis in VOSviewer with the *full counting* method applied to keywords, with a minimum threshold of 29 occurrences, which retained 70 of the 6,882 keywords (about 1%). Only this single threshold was applied, and no sensitivity analysis with other thresholds was conducted. The threshold of 29 retained 70 keywords (about 1% of the total), a number small enough to display as a readable map. Default VOSviewer settings were used, namely association strength normalization, a clustering resolution of 1.00, and a minimum cluster size of 1 (VOSviewer version 1.6.21). Performance analysis (number of publications per year, subject-area distribution, and country contributions) used the *Analyze Results* feature in Scopus. VOSviewer produced network, overlay, and density maps, which were interpreted descriptively based on node size, cluster color, link strength, color gradation, and density intensity.

III. RESULTS AND DISCUSSION

Results

Bibliometric analysis identifies research trends in GenAI in learning through publication productivity, keyword linkages, topic development, subject areas, and country contributions [23]. Publication trends, subject areas, and countries were derived from Scopus *Analyze Results*, whereas the network, overlay, and density maps were derived from VOSviewer. Each finding is first presented and then interpreted and compared with previous research in the Discussion. The results are presented in the order of the research questions, from the publication trend to the country contributions.

The first finding is the publication trend for 2023–2025, which shows the growth rate and the peak year (Figure 2). The trend was obtained from the Scopus *Analyze Results* feature for the 1,400 documents that passed screening. Figure 2 presents the number of documents in each year, which rose from 54 in 2023 to 998 in 2025. This shows how quickly the literature has expanded since ChatGPT appeared in late 2022.

The publication data cover 2023, 2024, and 2025, which are the three complete calendar years after ChatGPT appeared. The year 2026 was excluded because it was still incomplete when the data were downloaded, so its partial count would not be comparable with the full-year counts. Growth is expressed both as the ratio of documents between consecutive years and as the percentage increase from the previous year. The share of each year is calculated against the 1,400 documents that passed all screening stages.

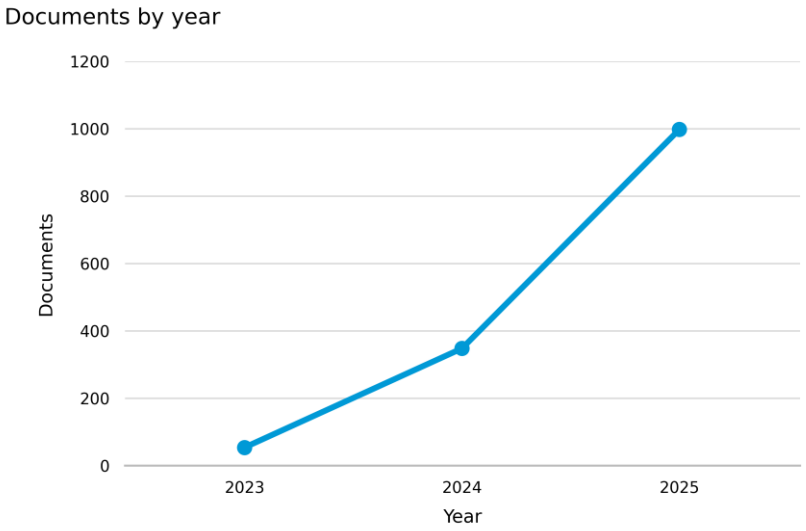


Figure 2. Distribution of publication trends, 2023–2025 (source: Scopus)

Figure 2 shows a sharp increase in GenAI publications in learning. In 2023, the first year after ChatGPT appeared, there were only 54 documents (about 3.9% of the total), indicating an early exploratory stage. The number rose to 348 documents in 2024 (an increase of about 544%, or 6.4-fold) and peaked at 998 documents in 2025 (about 187%, or 2.9-fold, above 2024). Documents from 2024 account for about 24.9% of the data, whereas documents from 2025 account for about 71%, so more than two-thirds of the literature comes from the most recent single year. This pattern indicates that GenAI has developed from an emerging topic into a major focus of educational research. However, because the growth was not compared with total Scopus output, it is read as a descriptive indication of rising scholarly attention rather than as evidence of a particular growth pattern.

The second finding comes from the keyword *co-occurrence* analysis in the network visualization (Figure 3). Nodes represent keywords, links indicate co-occurrence, and colors form interrelated thematic clusters. Of the 6,882 normalized keywords, 70 met the minimum threshold of 29 occurrences and formed four clusters. Figure 3 presents the resulting network map, and Table 1 summarizes the theme of each cluster.

The map was generated from the 6,882 normalized keywords by applying a minimum threshold of 29 occurrences, which retained 70 keywords or about 1% of the total. Node size indicates how often a keyword occurs, and the links between nodes indicate how often two keywords appear together in the same documents. Clustering used association strength normalization with a resolution of 1.00 and a minimum cluster size of 1, so cluster membership depends on these default settings. The four resulting clusters differ in size, with 23 keywords in the red cluster, 20 in the green cluster, 17 in the blue cluster, and 10 in the yellow cluster. Because overlapping labels are hidden by VOSviewer, only legible labels are listed in Table 1 and discussed in the text.

The network map is read in three steps. First, the largest nodes are identified as the most frequent keywords in the corpus. Second, the cluster colors are used to group keywords into thematic directions and to name each direction from its most frequent keywords. Third, the links between clusters are examined to see which themes are connected to one another. This reading is descriptive, and it does not test the relationships among keywords statistically. Keywords that lie between two clusters are of particular interest because they connect different themes. The thickness of a link between two keywords shows how often they are used together, so thick links point to stable associations in the literature. Weak links at the edge of the map show associations that appear

only occasionally. In this study the links are read as indications of association and not as evidence of causal relationships between the keywords or the topics they represent.

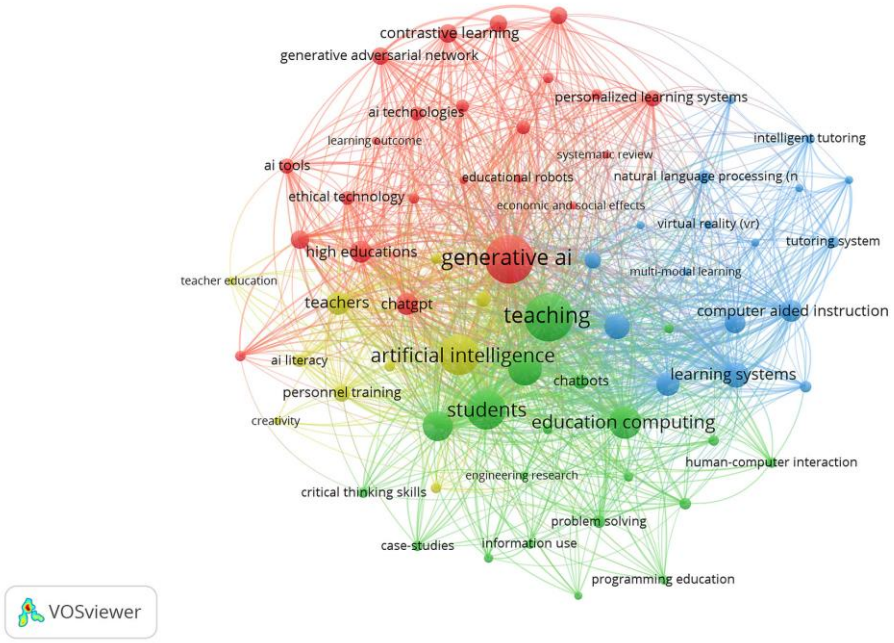


Figure 3. Network visualization of GenAI research in science learning (source: VOSviewer)

Table 1. Thematic clusters from the keyword co-occurrence analysis

Cluster	Theme	Representative keywords
Red	Generative AI and higher education	*generative ai, chatgpt, high educations, ai tools, ai technologies, ethical technology, personalized learning systems, contrastive learning, generative adversarial network, educational robots*
Green	Teaching, students, and computing education	*teaching, students, education computing, chatbots, problem solving, critical thinking skills, programming education, human-computer interaction, case-studies, information use*
Blue	Intelligent learning systems and computer-aided instruction	*computer aided instruction, learning systems, intelligent tutoring, tutoring system, natural language processing, virtual reality (vr), multi-modal learning*
Yellow	Artificial intelligence, teachers, and AI literacy	*artificial intelligence, teachers, teacher education, ai literacy, personnel training, creativity*

(Source: VOSviewer, processed. Minimum threshold of 29 occurrences; only keywords whose labels are legible on the map are listed. Theme names are interpretive labels based on the most frequent keywords in each cluster; some keywords, such as contrastive learning, fit their label only loosely. The clusters contain 23 (red), 20 (green), 17 (blue), and 10 (yellow) keywords)

Figure 3 displays four clusters in red, green, blue, and yellow (Table 1). The largest labeled nodes are "*generative ai*", "*teaching*", and "*artificial intelligence*" at the center of the map, connected to all four clusters; several other large nodes are unlabeled because VOSviewer hides overlapping labels. Because "generative ai" and "teaching" are also the search terms, their central position and large size are largely a result of the query, so the interpretation below focuses on the clusters and the keywords around these two nodes rather than on the nodes themselves. The dense links around the center show that the clusters are interrelated, so the field is multidisciplinary and connects perspectives on technology, teaching practice, the role of teachers, and learning systems. In physics learning, this structure opens opportunities to connect chatbots, tutoring systems, and natural language processing with specific physics concepts.

The third finding comes from the overlay visualization (Figure 4), which shows the temporal development of keywords according to their average year of appearance; color variation distinguishes earlier and more recent topics. The overlay map uses the same 70 keywords and four clusters as the network map, but node color now indicates average publication year instead of cluster membership. Figure 4 presents this map, and its color scale is read as a relative indication of which topics appeared earlier or later within 2023–2025. Because the period is short, the overlay is interpreted together with the network and density maps. The overlay map can therefore

suggest which topics were more prominent early in the period and which became more prominent later, but only as a relative indication. Differences in color are therefore compared with cluster membership to see which thematic directions are more recent.

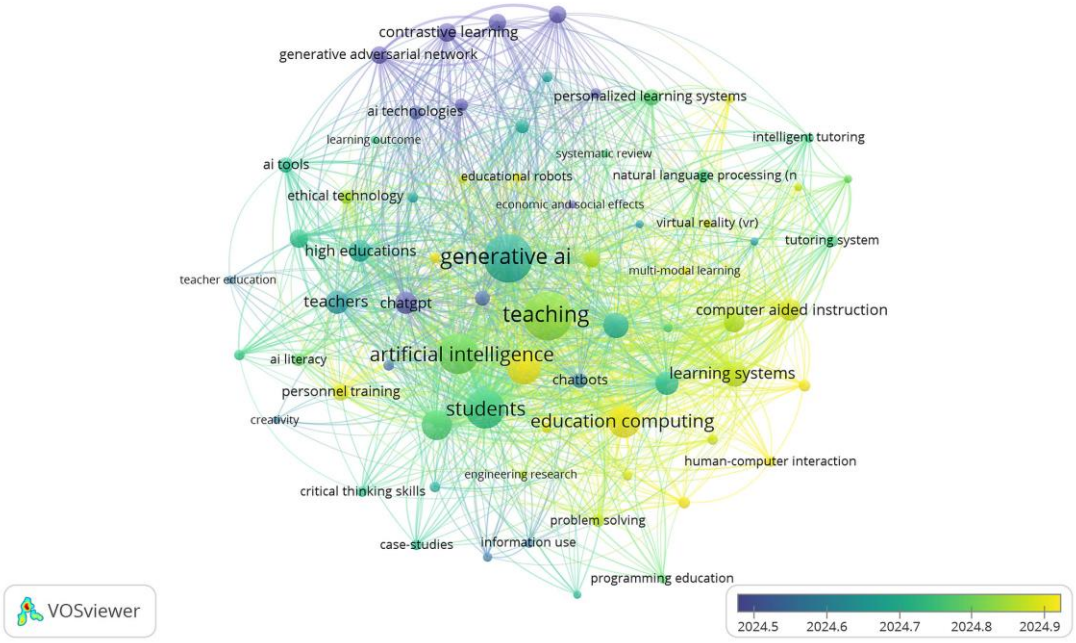


Figure 4. Overlay visualization (source: VOSviewer)

The color scale in Figure 4 ranges from an average publication year of 2024.5 (blue–purple) to 2024.9 (yellow), whereas the average years of all 70 keywords range only from 2024.29 to 2024.98, a span of less than one year, so differences among topics are relative and the overlay is read as a tentative tendency rather than a chronological order. The lowest average years (blue–purple) are found for *contrastive learning*, *generative adversarial network*, *ai technologies*, *chatgpt*, *economic and social effects*, *chatbots*, *teacher education*, and *information use; mid-range values (teal to green) for generative ai, high educations, ai tools, teachers, and students;* and the highest average years (yellowish green to yellow) for *teaching*, *computer aided instruction*, *learning systems*, *education computing*, *human-computer interaction*, *virtual reality (vr)*, *multi-modal learning*, *personnel training*, *problem solving*, *educational robots*, and *ethical technology*. Central keywords such as *generative ai* and *artificial intelligence* lie in the middle of the scale, consistent with their role as general concepts present throughout the period, whereas peripheral keywords such as *human-computer interaction*, *multi-modal learning*, and *educational robots* tend toward more recent years. Because no finer time resolution (for example, year-by-year maps) was used, these patterns are read together with the network and density maps and are not treated as evidence of a sharp shift.

The fourth finding comes from the density visualization (Figure 5). Areas of high density (yellow–green) indicate dominant themes, whereas areas of low density (blue–purple) indicate rarely studied topics as potential research gaps [27]. The density map is based on the same keyword data as the network and overlay maps, so it shows how frequently keywords occur in the corpus rather than how they are linked. Figure 5 presents this map, and its color gradation is used to identify dominant themes and possible research gaps.

In the density map, each keyword is given a color according to the density of its neighborhood, which depends on the number and weight of the keywords around it. Colors range from blue–purple for low density to yellow for high density, so bright areas mark keywords that are frequent and closely connected with others. The map therefore complements the network map by showing where research is concentrated and not only how keywords are linked. In this study, density is used to identify dominant themes and to suggest, tentatively, which topics remain rare. Because “generative ai” and “teaching” are the search terms, high density around them is expected and is not treated as a finding in itself.

Three questions guide the reading of Figure 5. The first asks which keywords lie in the highest-density areas, and the second asks which keywords lie in low-density areas. The third asks whether physics- or science-specific keywords appear as distinct nodes. The answers are compared with the network and overlay maps to

check whether the same pattern appears in all three views. This comparison across views reduces the risk of over-interpreting a single map. Possible gaps are treated as hypotheses to be tested with complementary searches and not as established conclusions.

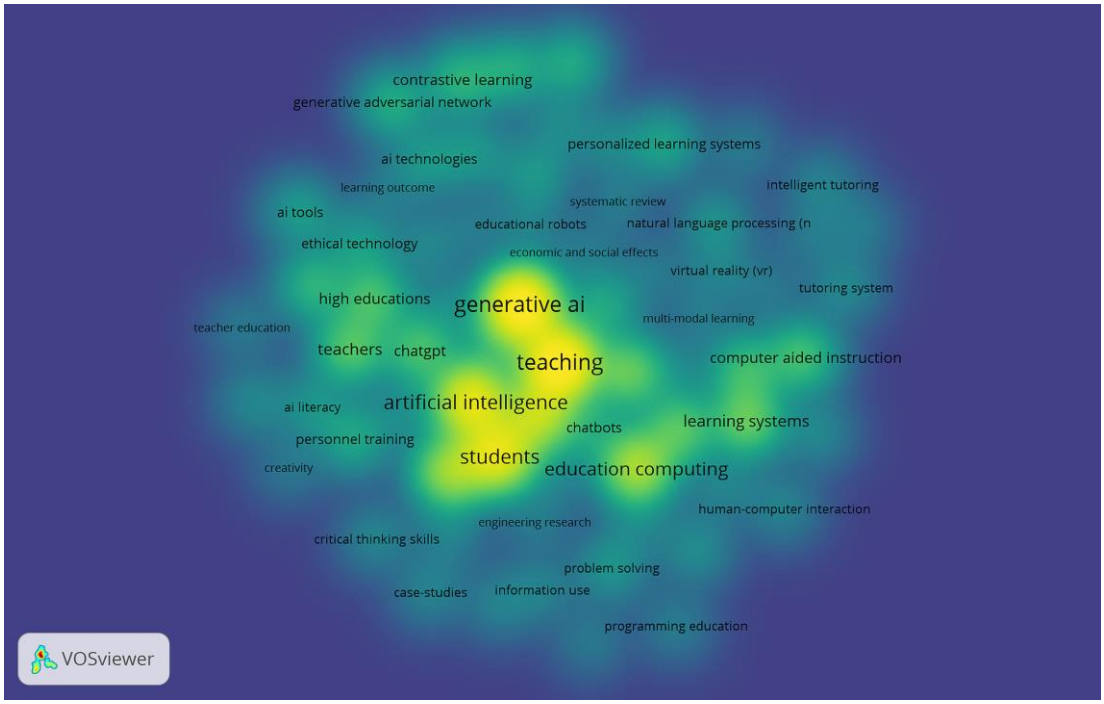


Figure 5. Density visualization (source: VOSviewer)

In Figure 5, the highest density (yellow) is centered on "*generative ai*", "*teaching*", "*artificial intelligence*", and "*students*", indicating the dominance of studies on the use of GenAI in teaching, although the high density of "generative ai" and "teaching" is largely a result of the query. A yellowish-green area appears around *education computing*, and green areas mark *learning systems*, *computer aided instruction*, *high educations*, and *teachers*. In contrast, darker regions with peripheral keywords such as *teacher education*, *creativity*, *human-computer interaction*, *tutoring system*, *case-studies*, and *information use* indicate less emphasized topics, while *ethical technology* and *ai literacy* appear in dimmer teal areas. This pattern points to possible research gaps in ethics, teacher readiness, and AI literacy, as well as in physics and science learning, which do not yet appear as distinct nodes. Low density only reflects low keyword frequency within this corpus and may also result from the search query and the subject-area filter, so these possible gaps should be tested with complementary searches (for example, "generative ai" AND ethics) in further work. This finding is consistent with the reading of the network and overlay maps, namely that the center of study lies in teaching with GenAI, whereas ethics, teachers, and literacy remain at the periphery.

The fifth finding comes from the subject-area distribution in Scopus (Figure 6), which shows the dominant field and the role of supporting fields. This distribution was obtained from the Analyze Results feature in Scopus and is presented as a pie chart, so the area of each slice represents the proportion of documents in a subject area. This information is important for identifying the disciplines that most often host research on GenAI in learning and for assessing how large a share the natural sciences, particularly physics, hold in the retrieved literature. In addition, the distribution of subject areas helps explain the disciplinary background of the thematic clusters seen in the earlier network map. Reading Figure 6 must also take into account the search strategy, which restricted subject areas at the screening stage, so the composition shown reflects the retrieved documents rather than the entire GenAI literature in Scopus. In this way, the dominance of a field, as well as fields that rarely appear, can be recognized directly from the slice sizes in the chart. The percentages in the chart are used to compare the roles of the main and supporting fields in the literature. This comparison is needed because the research topic lies at the intersection of technology and education, so the contribution of each field must be viewed proportionally.

The chart shows the ten largest subject areas, and the remaining areas are combined into the Other category. Each slice is labeled with its percentage, which makes it possible to compare the share of Computer Science with that of Physics and Astronomy directly. Physics and Astronomy is of special interest here because it represents the natural science discipline that this study seeks to connect with GenAI. The comparison is

therefore used later to judge how far physics-specific research appears in this literature. Slice sizes that are close to each other should be compared with the exact percentages rather than judged by eye. The reading of the chart is also limited by the subject-area filter used in the search strategy.

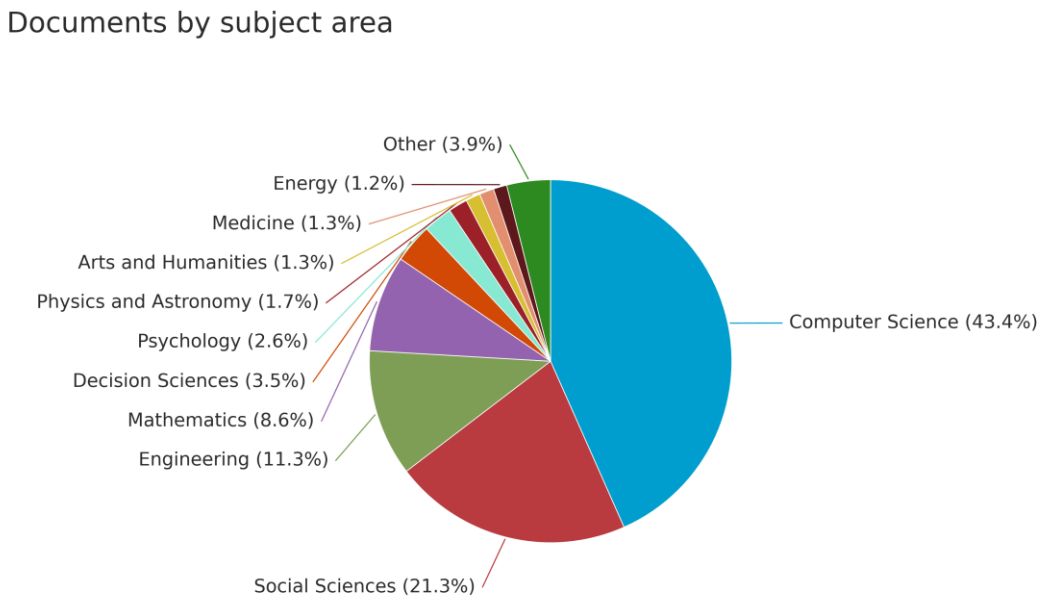


Figure 6. Subject-area distribution (source: Scopus)

According to Figure 6, Computer Science dominates with 1,380 documents (43.4%), followed by Social Sciences with 676 documents (21.3%), Engineering with 359 (11.3%), and Mathematics with 273 (8.6%). Other fields contribute less: Decision Sciences 112 (3.5%), Psychology 83 (2.6%), Physics and Astronomy 53 (1.7%), Arts and Humanities 42 (1.3%), Medicine 40 (1.3%), and Energy 38 (1.2%). Percentages were calculated against the total of subject-area classifications because one document can belong to more than one field. Earth and Planetary Sciences does not appear as a separate category and is therefore probably included in the Other category (3.9%), indicating that its contribution is relatively small. The dominance of Computer Science indicates a technological orientation, whereas Social Sciences reflects the pedagogical dimension.

The sixth finding comes from country contributions (Figure 7), which show the countries and regions with the largest roles in research on GenAI in learning in 2023–2025. Country contribution data were obtained from the Analyze Results feature in Scopus and are presented as a bar chart ranking countries and regions by number of documents. Counts are based on author affiliations, so a document written by authors from several countries can be counted for more than one country. This information is important for viewing the geographical distribution of research, identifying centers of knowledge production, and assessing opportunities for contributions from countries that are not yet prominent, including Indonesia. Figure 7 is read by considering country rankings and the gaps in document counts among countries, so the level of research concentration can be determined. This comparison also provides a basis for interpreting the position of developing countries on the map of GenAI research for science learning. Beyond document counts, country contributions can be assessed from the closeness of figures among countries in the middle to lower ranks, because small gaps indicate that countries in that group are relatively equal in strength. This way of reading helps separate countries that serve as research hubs from those that contribute moderately. The results are then related to previous studies to explain possible causes of the concentration of publications.

The ranking in Figure 7 is read in two ways. The first is the absolute number of documents, which shows which countries and regions produce the most research on GenAI in learning. The second is the gap between neighboring ranks, which shows whether the literature is spread evenly or concentrated in a few places. Because the authors of this study work in Indonesia, the position of Indonesia in this ranking is of particular interest for the implications discussed later.

The figure lists the ten countries and regions with the most documents, and Hong Kong is shown separately from China as a region. Counts follow author affiliations, so a collaboration among authors from several countries raises the totals of several countries at once. For this reason, the sum across countries is not expected to equal the 1,400 documents analyzed. The results are therefore read as the relative position of each country and

not as an exact partition of the corpus. Small differences in rank among the lower-ranked countries should therefore be interpreted with caution. These limits are kept in mind when country contributions are compared in the analysis.

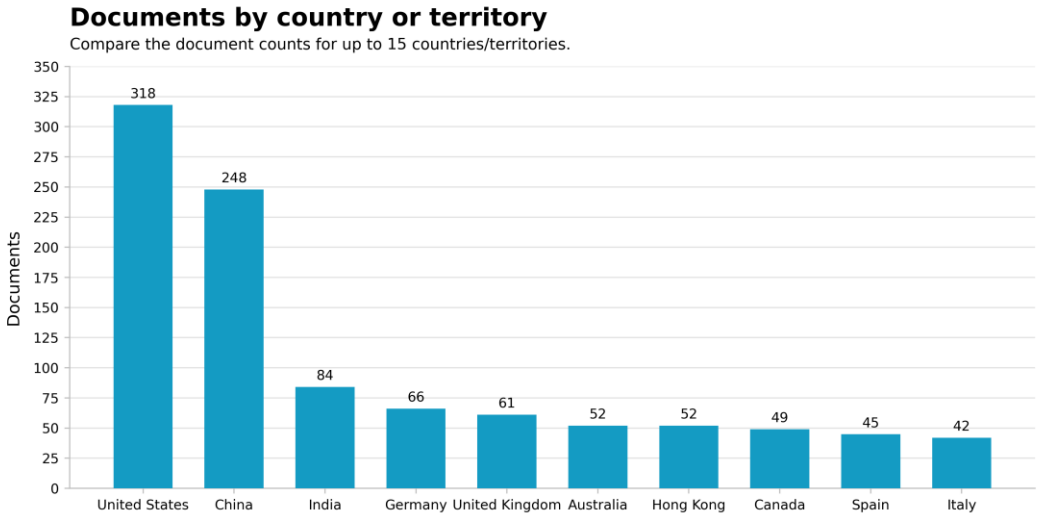


Figure 7. Distribution of publications by country (source: Scopus)

According to Figure 7, the top ten countries and regions are the United States (318 documents), China (248), India (84), Germany (66), the United Kingdom (61), Australia (52), Hong Kong (52), Canada (49), Spain (45), and Italy (42). The United States and China together contribute 566 documents, far above other countries. The United States alone has about 3.8 times as many documents as India, which ranks third. Of the top ten, only China and India are developing countries, and Indonesia is not included. The remaining eight countries and regions each contribute between 42 and 84 documents, so the distribution shows a clear gap between two leading countries and the rest of the group. This pattern suggests that research on GenAI in learning is concentrated in a few countries, so evidence from other educational contexts, including Indonesia, remains limited and offers room for further contributions.

Discussion

This subsection interprets the findings presented above and compares them with previous research. The discussion follows the order of the findings, beginning with the publication trend and ending with the contributions of countries. Where the corpus or the search design limits a comparison, the limitation is stated so that the interpretation is not overstated. The comparison also draws out implications for physics and science learning, which are developed further in the final subsections.

The sharp growth in publications is consistent with findings on the exponential growth of ChatGPT publications after its launch [30], although three yearly data points are not sufficient to test the form of growth statistically. The rising interest can also be linked to early findings that GenAI can be used to design lesson units, rubrics, and quizzes [8]. The increase from 54 to 998 documents may also partly reflect changing terminology and the growth of Scopus output as a whole, because growth relative to total Scopus output was not computed. Thus, the trend is best read as a descriptive indication of growing scholarly attention and not as evidence of a particular growth pattern.

In physics, early studies show two sides. ChatGPT's answers to physics questions were reported to be often contradictory and unreliable [31], whereas simple classroom activities using ChatGPT in physics were proposed to foster critical thinking [32]. ChatGPT's capabilities were tested on concept tests and graph interpretation [33], and ChatGPT nearly passed an introductory physics course while displaying novice-like preconceptions and errors [34]. More recent studies have turned to its impact on learning: a randomized controlled trial reported that an AI tutor built on pedagogical principles produced higher physics learning gains in less time than an active-learning class. Thus, GenAI has become a major agenda in educational research, whereas physics-specific studies are only beginning to take shape.

This multi-cluster structure is consistent with earlier mapping studies [35] and similar to the clusters on AI in science education [19]. The red and green clusters, which contain *generative ai*, *chatgpt*, *ai tools*, and *chatbots*, can be compared with the thematic structures reported in mappings of GenAI in education [20] and in higher education; a one-to-one comparison is not possible because those studies use different scopes and corpora. The blue cluster, which contains *intelligent tutoring*, *tutoring system*, and *learning systems*, reflects the role of AI as a learning mediator [1] and continues the tradition of intelligent tutoring systems, which a meta-analysis found to have a median effect of 0.66 standard deviations above conventional instruction [36]. Feedback quality in such systems matters because feedback is among the strongest influences on learning outcomes, although its impact can be positive or negative depending on its quality [37].

The yellow cluster, which contains *teachers*, *ai literacy*, *personnel training*, and *teacher education*, is consistent with the assertion that using GenAI requires the ability to evaluate the accuracy of information and to use it responsibly. An exploratory review of 30 articles proposed four aspects of AI literacy relevant to teacher training: know and understand, use and apply, evaluate and create, and ethical issues [38]. The red cluster, which contains *high educations* and *ethical technology*, is consistent with the emphasis on academic integrity and the difficulty of detecting GenAI use as a major issue in higher education [39], as well as with mappings of GenAI in higher education. Together, the yellow and red clusters show that attention to teachers, literacy, and ethics runs alongside the technology-oriented keywords in the field.

The green cluster, centered on *teaching*, *students*, and *education computing* and including *programming education* and *problem solving*, shows that GenAI research largely originates from computing teaching practice, in line with the dominance of Computer Science (Figure 6). *The diversity of clusters is consistent with* the very high heterogeneity ($I^2 = 96.32\%$) in the meta-analysis, which concluded that GenAI is promising when used to enrich, not replace, learning activities [16]. Science-specific keywords such as physics do not appear among the labels displayed, which may also relate to the threshold of 29 occurrences. Thus, GenAI research is developing from four complementary directions, namely generative AI and the higher education context, teaching practice and computing education, tutoring systems and computer-aided instruction, and the roles and literacy of teachers, whereas the specificity of physics and science is not yet clearly visible.

A similar tendency from tool adoption toward learning processes and assessment has been reported for GenAI in higher education [21]. The emergence of *computer aided instruction*, *learning systems*, and *human-computer interaction* as the most recent topics may indicate a shift of attention from the ability to answer toward the quality of learning interaction. In physics, in an augmented-reality quantum cryptography laboratory experiment, ChatGPT-based feedback significantly improved students' learning outcomes and influenced their eye-tracking data. However, AI assistance does not automatically guarantee learning. In a field experiment with nearly a thousand high school students, unguarded access to GPT-4 led students to use it as a shortcut and to perform worse once access was removed (a 17% drop in grades in the GPT Base group), whereas a tutor with pedagogical safeguards largely mitigated this negative effect. That study was conducted in mathematics, but it is relevant to physics because it is consistent with findings that ChatGPT's performance on physics tasks involving graphs is not yet fully reliable [40].

The emergence of *multi-modal learning* signals interest in multimodal data, which promises a more complete picture of the learning process while also presenting its own challenges [41]. In contrast, *chatgpt*, *generative adversarial network*, and *contrastive learning* in the early-topic group show that the initial phase of research was dominated by the exploration of tools and techniques. Research appears to be moving from questions about what GenAI can do toward how GenAI supports learning processes, although this sequence is only a relative tendency because the range of average years is only 2024.5–2024.9. These tendencies are only suggestive, so they should be confirmed with year-by-year maps in further work.

The low density of *ethical technology*, *ai literacy*, and *teacher education* deserves attention; *ethical technology* is among the more recent topics on the overlay map but remains thin on the density map. Ethical intent alone is not sufficient in AI for education and must be accompanied by explicit consideration of fairness, accountability, transparency, bias, autonomy, agency, and inclusion [42]. Long before GenAI, the near absence of critical reflection on the challenges and risks of AI in higher education was already documented [43], and a similar gap may persist in the GenAI literature. Responsible implementation and AI literacy are prerequisites for ideal GenAI use [2], which relates to the challenges of hallucination, bias, and teacher readiness that have been noted [14]. It likewise encourages regulation, data protection, and capacity building for educators in the use of GenAI [44].

In physics learning, the need for critical evaluation is also real. All ChatGPT answers to mechanics comprehension questions were found to be incorrect, inaccurate, incomplete, or misleading, while students'

ratings distinguished the reference solutions from ChatGPT's answers only on questions of low and medium difficulty. In Indonesia, a study of 56 eleventh-grade senior high school students in Jombang found that students tended to accept GenAI answers quickly and only a small proportion were accustomed to questioning, verifying, and evaluating them. Ethical and reflective use correlated moderately and positively with critical thinking in physics ($\rho = 0.525$), and teacher supervision and ethical safeguards were recommended. These findings indicate that ethics, policy, and teacher training are prerequisites for accurate GenAI implementation in physics and science learning, and therefore merit consideration as possible research gaps.

The dominance of Computer Science is also influenced by the search strategy, which restricted the subject areas to three fields; Social Sciences and Engineering appear because documents can be classified into more than one field. Mapping studies of AI in science education focus on the science education domain, so their subject-area composition is not directly comparable [17]. The strong technological orientation needs to be read together with the observation that AI research in higher education remains weak in critical reflection and in linkage with pedagogical perspectives [43]. Physics and Astronomy contains only 53 documents (1.7%) according to the Scopus classification, which indicates that studies of GenAI for physics learning remain limited, although this figure should be read with caution because physics education documents classified only under other fields, such as Social Sciences, were not captured by the subject-area filter. A mapping of AI in physics education recommended evaluating the accuracy of AI models across various physics topics and developing fair and transparent AI-based assessment [18]. Because its scope is AI in general, the present study complements it with a GenAI-specific picture suggesting that the share of physics in this literature is small, within the limits of the subject-area filter.

The dominance of the United States and China, followed by India, Germany, and the United Kingdom, shows that research is concentrated in a few countries, which is probably related to large research capacity and resonates with discussions of power dynamics among countries in GenAI research [22]. Indonesia's absence from the top ten leaves room for contributions from science learning contexts in developing countries. Empirical evidence from Indonesia is beginning to emerge, for example on patterns of GenAI use among senior high school students in physics learning [12], but it is small in scale (56 students at one school), so its generalizability must be treated with caution. This space matters because implementing GenAI requires policy and capacity building at the national level, so evidence from dominant countries does not automatically apply in other contexts without being tested in local settings.

Implications and directions for future research. In response to the seventh research question, this mapping points to three possible research gaps and opportunities for science and physics education researchers, which are hypotheses to be tested with broader searches rather than data-based conclusions. First, research on GenAI for specific physics concepts, because Physics and Astronomy accounts for only 1.7% in the Scopus classification and physics-specific keywords do not yet appear as distinct nodes. Such research can extend earlier evidence [6,34,45] to other topics, educational levels, and contexts. Second, studies of ethics, AI literacy, and teacher readiness, which have low density, including training of physics teachers and prospective teachers to critically evaluate GenAI outputs, as the need was shown by earlier findings [11] and supported by guidance on the responsible use of GenAI in education [44]. Third, the implementation of GenAI for tutoring systems, computer-aided instruction, and human–computer interaction as the most recent topics, with designs that account for the risk of dependence [10] and the benefits of feedback [7]. Contributions from the contexts of developing countries, including Indonesia, may also be needed.

Limitations. The mapping results depend on data quality and on parameters such as the threshold and normalization [46]. The threshold of 29 occurrences filters out low-frequency keywords, so specialized topics, including physics topics, may not appear on the map. Only this single threshold was used, so the stability of the clusters under other thresholds was not tested. This study is limited to Scopus, the years 2023–2025, and three selected subject areas, so documents classified only under other fields, such as Social Sciences, were not captured. Social Sciences is where most education and science-education journals are indexed, so this restriction is a serious limitation of the design. 2026 was excluded because it was incomplete at the time of download, so the most recent trend is not captured. Scopus journal coverage differs from that of Web of Science and Dimensions [47], so results may differ in other databases. The keyword pair "*generative ai*" and "*teaching*" may also miss documents that use other terms, such as "ChatGPT", "large language model", "GenAI", "generative artificial intelligence", "learning", or "education". Because "generative ai" became common mainly after 2023, the growth from 54 to 998 documents may partly reflect changing terminology and the growth of Scopus output as a whole; growth relative to total Scopus output was not computed. Because "generative ai" and "teaching" are also the search terms, their central position and density in the maps largely reflect the query. In addition, because

the scope of science was approached through subject areas rather than science keywords, the search may capture documents that are not specific to science learning and miss science documents that lack the word “teaching”. Because of this restriction, almost all documents (1,380 of 1,400) are labeled Computer Science, so the thematic clusters, particularly computing education, reflect learning contexts in the selected subject areas rather than all science learning. Finally, the analysis was limited to publication performance by year, subject area, and country and to keyword linkages; productive sources, authors, and institutions were not analyzed, and the quality, citations, and empirical results of each document were not examined, so these findings describe the structure and direction of topics, not the effectiveness of GenAI. The gaps identified here are possible gaps inferred from keyword frequency and were not tested with complementary searches.

IV. CONCLUSION

Research on GenAI in learning increased sharply from 54 documents in 2023 to 998 documents in 2025. The network analysis showed four thematic clusters centered on “*generative ai*”, “*teaching*”, and “*artificial intelligence*”, while the overlay, within a very narrow time range (2024.5–2024.9), suggested only a tentative tendency from early topics such as *chatgpt* and generative techniques toward learning systems, computer-aided instruction, and computing education. The density visualization showed a concentration on studies of generative AI in teaching, with technology ethics, AI literacy, and teacher education as relatively less prominent topics. Computer Science dominates (43.4% of all subject-area classifications), whereas Physics and Astronomy accounts for only 1.7% according to the Scopus classification, and the United States and China are the largest contributors. Because the corpus is almost entirely Computer Science (1,380 of 1,400 documents), these findings describe GenAI in computing-dominated teaching contexts and should not be generalized to science learning as a whole. Within these limits, the mapping points to possible gaps in physics-specific topics, teacher readiness, and ethical use, which should be tested in future studies that widen the query with science terms and include documents classified under Social Sciences.

ACKNOWLEDGMENT

The authors sincerely thank all parties who provided support, guidance, and valuable input during the research and the writing of this article. All of this assistance and encouragement was very meaningful for completing the research and refining the manuscript. Special thanks are extended to the lecturer who reviewed the manuscript and gave detailed suggestions that improved its quality. The authors also thank colleagues and fellow researchers who took part in discussions on the research design, the data analysis, and the interpretation of the results.

REFERENCES

[1] W. Xu and F. Ouyang, “The application of AI technologies in STEM education: a systematic review from 2011 to 2021,” *Int. J. STEM Educ.*, vol. 9, no. 1, p. 59, 2022, doi: 10.1186/s40594-022-00377-5.

[2] E. Kasneci *et al.*, “ChatGPT for good? On opportunities and challenges of large language models for education,” *Learn. Individ. Differ.*, vol. 103, p. 102274, 2023, doi: https://doi.org/10.1016/j.lindif.2023.102274.

[3] A. Yusuf, N. Pervin, M. Román-González, and N. M. Noor, “Generative AI in education and research: A systematic mapping review,” *Rev. Educ.*, vol. 12, no. 2, p. e3489, Aug. 2024, doi: https://doi.org/10.1002/rev3.3489.

[4] Kamaliatul hasanah, E. V. Aulia, and M. A. Mahdiannur, “Systematic Review : Penerapan Kecerdasan Buatan (AI) dalam Pembelajaran IPA Meningkatkan Pemahaman,” *Edukimbiosis J. Pendidik. IPA*, vol. 4, no. 1, pp. 26–34, 2025, doi: 10.35905/edukimbiosis.v4i1.15118.

[5] A. Asrizal, N. Annisa, F. Festiyed, H. Ashel, and R. Amnah, “STEM-integrated physics digital teaching material to develop conceptual understanding and new literacy of students,” *Eurasia J. Math. Sci. Technol. Educ.*, vol. 19, p. em2289, Jul. 2023, doi: 10.29333/ejmste/13275.

[6] G. Kestin, K. Miller, A. Klales, T. Milbourne, and G. Ponti, *AI Tutoring Outperforms Active Learning*. 2024. doi: 10.21203/rs.3.rs-4243877/v1.

[7] A. Coban, D. Dzsotjan, S. Küchemann, J. Durst, J. Kuhn, and C. Hoyer, “AI support meets AR visualization for Alice and Bob: personalized learning based on individual ChatGPT feedback in an AR quantum cryptography experiment for physics lab courses,” *EPJ Quantum Technol.*, vol. 12, no. 1, p. 15,

2025, doi: 10.1140/epjqt/s40507-025-00310-z.
[8] G. Cooper, "Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence," *J. Sci. Educ. Technol.*, vol. 32, no. 3, pp. 444–452, 2023, doi: 10.1007/s10956-023-10039-y.
[9] S. Küchemann *et al.*, "Can ChatGPT support prospective teachers in physics task development?," *Phys. Rev. Phys. Educ. Res.*, vol. 19, Sep. 2023, doi: 10.1103/PhysRevPhysEducRes.19.020128.
[10] H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman, "Generative AI without guardrails can harm learning: Evidence from high school mathematics," *Proc. Natl. Acad. Sci.*, vol. 122, no. 26, p. e2422633122, Jul. 2025, doi: 10.1073/pnas.2422633122.
[11] M. N. Dahlkemper, S. Z. Lahme, and P. Klein, "How do physics students evaluate artificial intelligence responses on comprehension questions? A study on the perceived scientific accuracy and linguistic quality of ChatGPT," *Phys. Rev. Phys. Educ. Res.*, vol. 19, no. 1, p. 10142, Jun. 2023, doi: 10.1103/PhysRevPhysEducRes.19.010142.
[12] A. Zakaria, B. Jatmiko, B. K. Prahani, and C. F. Wijaksono, "Identification of Generative AI Usage Profiles and Implications for Critical Thinking Skills in Physics Learning," *JPPS (Jurnal Penelit. Pendidik. Sains)*, vol. 15, no. 2 SE-Issue & Trend of Digital Technology in Science Education, pp. 83–93, May 2026, doi: 10.26740/jpps.v15n2.p83-93.
[13] F. Almasri, "Exploring the Impact of Artificial Intelligence in Teaching and Learning of Science: A Systematic Review of Empirical Research," *Res. Sci. Educ.*, vol. 54, no. 5, pp. 977–997, 2024, doi: 10.1007/s11165-024-10176-3.
[14] A. K. Erümit and R. Ö. Sarıalioğlu, "Artificial intelligence in science and chemistry education: a systematic review," *Discov. Educ.*, vol. 4, no. 1, p. 178, 2025, doi: 10.1007/s44217-025-00622-3.
[15] K. K. C. Cheung, A. Zerouali, J. Koenen, and S. Erduran, "Do generative artificial intelligence (GenAI) and science education mix? A systematic review of the literature," *Stud. Sci. Educ.*, pp. 1–29, Oct. 2025, doi: 10.1080/03057267.2025.2578091.
[16] C. Boolzen, J. Kuhn, S. Flegr, E.-M. Rott, N. Stausberg, and S. Küchemann, "Evidence of impact and interpretational limits of generative AI in STEM education: a systematic review and meta-analysis on cognitive learning outcomes," *Artif. Intell. Rev.*, vol. 59, no. 10, p. 223, 2026, doi: 10.1007/s10462-026-11665-9.
[17] H. Genc and N. Koçak, "Bibliometric Analysis of Studies on the Artificial Intelligence in Science Education with VOSviewer," *J. Educ. Sci. Environ. Heal.*, pp. 183–195, Oct. 2024, doi: 10.55549/jeseh.756.
[18] S. Nurjanah, N. Martaputri, Z. Zamzami, I. Suardi, and H. Hulu, "Artificial Intelligence in Physics Education Research in Two Decades: A Bibliometric Study from Scopus Database," *J. Pendidik. Fis.*, vol. 12, pp. 53–86, Apr. 2024, doi: 10.26618/jpf.v12i2.14745.
[19] K. Kesgin, "Bibliometric analysis and predictive modeling map the role of artificial intelligence in science education from research trends to classroom integration," *Discov. Educ.*, vol. 4, no. 1, p. 464, 2025, doi: 10.1007/s44217-025-00906-8.
[20] J. Liu *et al.*, "A bibliometric analysis of generative AI in education: current status and development," *Asia Pacific J. Educ.*, vol. 44, pp. 156–175, Jan. 2024, doi: 10.1080/02188791.2024.2305170.
[21] S. Wang, Y. Zhong, and W. Li, "Bibliometric mapping of generative artificial intelligence research in higher education using VOSviewer and CiteSpace," *Discov. Artif. Intell.*, vol. 6, no. 1, p. 839, 2026, doi: 10.1007/s44163-026-01985-8.
[22] K. Dai, Y. Liu, and X. Zhang, "Generative AI in higher education: A bibliometric review of emerging trends, power dynamics, and global research landscapes," *Comput. Educ. Artif. Intell.*, vol. 10, p. 100544, 2026, doi: https://doi.org/10.1016/j.caeai.2026.100544.
[23] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, pp. 285–296, 2021, doi: https://doi.org/10.1016/j.jbusres.2021.04.070.
[24] F. Festiyed, Y. Tanjung, and M. A. Fadillah, "ChatGPT in Science Education: A Visualization Analysis of Trends and Future Directions," *JOIV Int. J. Informatics Vis.*, vol. 8, pp. 1614–1624, Dec. 2024, doi: 10.62527/joiv.8.3-2.2987.
[25] C. Liang, Y. Qian, Y. Lin, and Y. Wang, "Physics Education under the Application of Artificial Intelligence : Bibliometric Analysis Based on Web of Science Core Library (2021-2025)," vol. 1, pp. 1–16, 2026, doi: https://doi.org/10.48550.
[26] O. Öztürk, R. Kocaman, and D. Kanbach, "How to design bibliometric research: an overview and a framework proposal," *Rev. Manag. Sci.*, vol. 18, pp. 3333–3361, Mar. 2024, doi: 10.1007/s11846-024-00738-0.
[27] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric

mapping," *Scientometrics*, vol. 84, pp. 523–538, Dec. 2009, doi: 10.1007/s11192-009-0146-3.
[28] J. Baas, M. Schotten, A. Plume, G. Côté, and R. Karimi, "Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies," *Quant. Sci. Stud.*, vol. 1, pp. 377–386, Jan. 2020, doi: 10.1162/qss_a_00019.
[29] I. Zupic and T. Čater, "Bibliometric Methods in Management and Organization," *Organ. Res. Methods*, vol. 18, pp. 429–472, Dec. 2014, doi: 10.1177/1094428114562629.
[30] F. Farhat *et al.*, "The scholarly footprint of ChatGPT: a bibliometric analysis of the early outbreak phase," *Front. Artif. Intell.*, vol. 6, p. 1270749, Jan. 2024, doi: 10.3389/frai.2023.1270749.
[31] B. Gregorcic and A.-M. Pendrill, "ChatGPT and the frustrated Socrates," *Phys. Educ.*, vol. 58, p. 35021, Mar. 2023, doi: 10.1088/1361-6552/acc299.
[32] P. Bitzenbauer, "ChatGPT in physics education : A pilot study on easy-to- implement activities," vol. 15, no. 3, 2023.
[33] G. Polverini and B. Gregorcic, "Performance of ChatGPT on the test of understanding graphs in kinematics," *Phys. Rev. Phys. Educ. Res.*, vol. 20, p. 10109, Feb. 2024, doi: 10.1103/PhysRevPhysEducRes.20.010109.
[34] G. Kortemeyer, *Could an Artificial-Intelligence agent pass an introductory physics course?* 2023. doi: 10.48550/arXiv.2301.12127.
[35] G. Lee, M. Yun, X. Zhai, and K. Crippen, "Artificial Intelligence in Science Education Research: Current States and Challenges," *J. Sci. Educ. Technol.*, vol. 35, no. 1, pp. 110–127, 2026, doi: 10.1007/s10956-025-10239-8.
[36] J. Kulik and J. D. Fletcher, "Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review," *Rev. Educ. Res.*, vol. 86, pp. 42–78, Mar. 2016, doi: 10.3102/0034654315581420.
[37] J. Hattie and H. Timperley, "The Power of Feedback," *Rev. Educ. Res.*, vol. 77, no. 1, pp. 81–112, Mar. 2007, doi: 10.3102/003465430298487.
[38] D. T. K. Ng, J. Leung, S. Chu, and M. Shen, "Conceptualizing AI literacy: An exploratory review," *Comput. Educ. Artif. Intell.*, vol. 2, p. 100041, Nov. 2021, doi: 10.1016/j.caeai.2021.100041.
[39] D. Cotton, P. Cotton, and R. Shipway, "Chatting and cheating: Ensuring academic integrity in the era of ChatGPT," *Innov. Educ. Teach. Int.*, vol. 61, Mar. 2023, doi: 10.1080/14703297.2023.2190148.
[40] G. Polverini and B. Gregorcic, *Evaluating vision-capable chatbots in interpreting kinematics graphs: a comparative study of free and subscription-based models*. 2024. doi: 10.48550/arXiv.2406.14685.
[41] M. Cukurova, M. Giannakos, and R. Martinez-Maldonado, "The promise and challenges of multimodal learning analytics," *Br. J. Educ. Technol.*, vol. 51, pp. 1–9, Jul. 2020, doi: 10.1111/bjet.13015.
[42] W. Holmes *et al.*, "Ethics of AI in Education: Towards a Community-Wide Framework," *Int. J. Artif. Intell. Educ.*, vol. 32, pp. 504–526, Apr. 2021, doi: 10.1007/s40593-021-00239-1.
[43] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?," *Int. J. Educ. Technol. High. Educ.*, vol. 16, no. 1, p. 39, 2019, doi: 10.1186/s41239-019-0171-0.
[44] S. Boonlue, "GUIDANCE FOR GENERATIVE AI IN EDUCATION AND RESEARCH" FOR TEACHERS," *J. Ind. Educ.*, vol. 23, Aug. 2024, doi: 10.55003/JIE.23202.
[45] F. Kieser, P. Wulff, J. Kuhn, and S. Küchemann, *Educational data augmentation in physics education research using ChatGPT*. 2023. doi: 10.48550/arXiv.2307.14475.
[46] J. Moral-Munoz, E. Herrera-Viedma, A. Espejo, and M. Cobo, "Software tools for conducting bibliometric analysis in science: An up-to-date review," *El Prof. la Inf.*, vol. 29, Jan. 2020, doi: 10.3145/epi.2020.ene.03.
[47] V. K. Singh, P. Singh, M. Karmakar, J. Leta, and P. Mayr, "The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis," *Scientometrics*, vol. 126, no. 6, pp. 5113–5142, 2021, doi: 10.1007/s11192-021-03948-5.