

Artificial Intelligence in School Mathematics Education: A Bibliometric Trend Analysis**Yog Pragya¹ ; Sangita Kumari² ; Dr. Vinod Kumar Singh³****DOI: <http://doi.org/10.5281/zenodo.23093835>****Review: 30/08/2026****Acceptance: 04/09/2026****Publication: 02/10/2026**

Abstract: Artificial intelligence (AI) is now used widely to support mathematics learning in schools, but the growth of this research field has not been documented in a clear way. This study used bibliometric analysis to describe the trend, the main contributors, and the main themes of research on AI in school mathematics education. A Scopus search for 2016 to 2026 gave 294 English-language documents. The results were analyzed by year, country, affiliation, author, funding sponsor, and a supporting keyword analysis. From 2016 to 2023, output increased gradually to 47 documents; after that, it increased significantly to 116 documents in 2025 and 98 in 2026. The United States (60 documents) and China (47) dominated the field, followed by Hong Kong, Germany, and Indonesia. No single institution or author dominated. The National Science Foundation and the National Natural Science Foundation of China were the largest funders. Research on ChatGPT and large language models appeared after 2023 and now appears in nearly half of all documents. The findings can help researchers and policy makers plan future work.

Keywords: artificial intelligence, mathematics education, school education, generative AI, intelligent tutoring systems, bibliometric analysis

Introduction

Mathematics is an essential subject for schools across most countries; however, mathematics is difficult for many students and also very challenging for many teachers to provide individual attention. Artificial intelligence (AI) offers new ways to support both. Intelligent tutoring systems, adaptive learning platforms, and, more recently, generative AI tools can give hints, check answers, and adjust tasks to the level of the student (Hwang et al., 2020). Earlier meta-analyses found that intelligent tutoring systems can improve learning outcomes (Kulik & Fletcher, 2016; Ma et al., 2014), and large field trials have tested such systems in real schools (Pane et al., 2014).

Since the public release of ChatGPT in late 2022, the discussion has changed quickly. Researchers and teachers now ask how large language models can support mathematical reasoning and problem solving, and what risks these tools bring. International organizations have also started to give guidance on the use of generative AI in education (Miao & Holmes, 2023). As a result, many new studies have appeared in a short time, and it is hard for readers to see the overall picture of this research.

Bibliometric analysis is a useful way to build this picture. It uses publication data to describe how a field grows, who contributes to it, and which topics receive attention (Donthu et al., 2021; Zupic & Čater, 2015). Several reviews of AI in education exist (Chen et al., 2020; Zawacki-Richter et al., 2019; Zhai et al., 2021), and Hwang and Tu (2021) conducted research on AI in mathematics education. However, these works were published before the recent rise of generative AI, and they did not focus on school-level mathematics. This gap is important

¹ Senior Research Scholar, Faculty of Education, Banaras Hindu University, Varanasi, India

² Junior Research Scholar, Faculty of Education, Banaras Hindu University, Varanasi, India

³ Assistant Professor, Faculty of Education, Banaras Hindu University, Varanasi, India

because school teachers and policy makers need current evidence about where the research is growing and who is producing it.

The purpose of this study is to describe the trend of research on AI in school mathematics education from 2016 to 2026. The study answers five research questions:

- RQ1. How has the number of publications changed from 2016 to 2026?
- RQ2. Which countries and territories contribute most to this research?
- RQ3. Which institutions and authors are the most productive?
- RQ4. Which funding sponsors support this research?
- RQ5. Which research themes appear most often, and how have they changed over time?

Method

Research Design and Data Source

This study used a bibliometric design with descriptive analysis. Data were taken from Scopus, a large database of peer-reviewed literature that is widely used in bibliometric research because its records are curated and of high quality (Baas et al., 2020).

Search Strategy

The search combined three groups of terms (see Table 1) in the title, abstract, and keyword fields. Terms inside a group were joined with OR, and the three groups were joined with AND. The search was limited to documents published from 2016 to 2026 and written in English. Review documents were excluded. The search returned 294 documents, and all of them were kept for analysis.

Table 1: Search Strategy Used in Scopus

Concept	Search terms (title, abstract, keywords)	Joined by
Artificial intelligence	“artificial intelligence” OR “generative AI” OR “ChatGPT” OR “large language model” OR “intelligent tutoring system”	AND
Mathematics	“mathematics education” OR “mathematics learning” OR “mathematical reasoning” OR “mathematical problem solving”	AND
School level	“school” OR “K-12” OR “primary school” OR “middle school” OR “secondary school”	AND
Filters	Publication year 2016–2026; language: English; document type “review” excluded	—

Data Analysis

The analysis followed the steps described by Donthu et al. (2021). First, the document counts were obtained directly from the Scopus database using its “Analyze search results” tool. Second, percentages were calculated using 294 as the base. For countries, affiliations, authors, and funders, one document can be counted more than once when it has several contributors, so these percentages do not add up to 100. Third, the years were divided into three phases (2016–2019, 2020–2023, and 2024–2026) to compare growth, and the compound annual growth rate was calculated for 2020 to 2025. Finally, for RQ5, the author keywords were counted after joining

spelling variants (for example, “LLM” and “large language models”), and key terms were searched in titles, abstracts, and keywords to show how themes changed by year. The counts for year, document type, and country were cross-checked within Scopus to confirm that they were consistent.

Results

RQ1: Publication Trend, 2016–2026

Figure 1 and Table 2 show the number of documents per year. Research output was very small at the starting years. Only 12 documents (4.1% of the total) were published from 2016 to 2019, an average of three per year. In the second phase (2020–2023), output rose to 35 documents, or about nine per year, and it stayed at 10 documents in both 2022 and 2023.

Documents by year

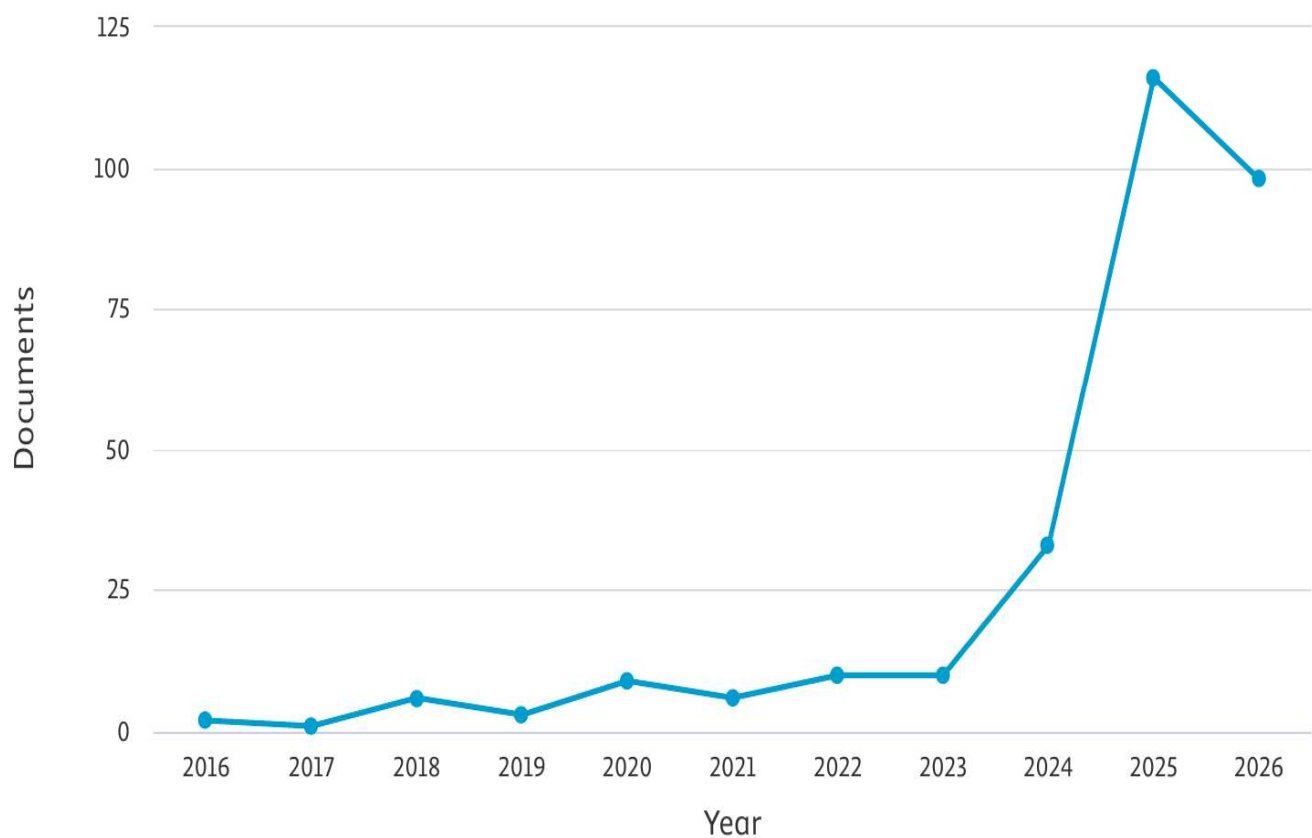


Figure 1: Number of Documents on AI in School Mathematics Education by Year, 2016–2026

The third phase (2024–2026) shows a very strong rise. Output grew from 10 documents in 2023 to 33 in 2024, an increase of 230%, and then to 116 in 2025, an increase of about 252%. The year 2025 alone produced more documents than the nine years from 2016 to 2024 together (80 documents). By August 2026, 98 documents had already been indexed, which is also higher than the total for 2016 to 2024. Because 2026 is incomplete (up to August 2026), the drop from 2025 to 2026 should not be read as a decline. Overall, 247 documents (84.0%) were published in the last three years. The compound annual growth rate from 2020 (9 documents) to 2025 (116 documents) was about 67%.

Table 2: Documents by Year and Phase

Year	Documents	% of total	Cumulative	Phase
2016	2	0.7	2	1 (2016–2019)
2017	1	0.3	3	
2018	6	2.0	9	
2019	3	1.0	12	
2020	9	3.1	21	2 (2020–2023)
2021	6	2.0	27	
2022	10	3.4	37	
2023	10	3.4	47	
2024	33	11.2	80	3 (2024–2026)
2025	116	39.5	196	
2026*	98	33.3	294	

Note. *2026 is a partial year (data up to August 2026).

RQ2: Contributing Countries and Territories

Figure 2 shows the ten leading countries and territories. The United States was first with 60 documents (20.4% of all documents), followed by China with 47 documents (16.0%). Hong Kong, which Scopus lists as a separate territory, was third with 18 documents (6.1%). Germany and Indonesia each had 15 documents (5.1%), Turkey had 11, India and Malaysia each had 9, and Italy had 8. Another 11 documents had no country information (“Undefined”).

There are two main points. First, the gap between the two leader countries and the rest is large: the United States and China together account for 107 country counts, while the third-ranked territory has 18. Second, five of the ten places are held by Asian countries or territories (China, Hong Kong, Indonesia, India, and Malaysia), and Europe is present through Germany and Italy. This shows that the topic is international and not limited to the traditional research centers.

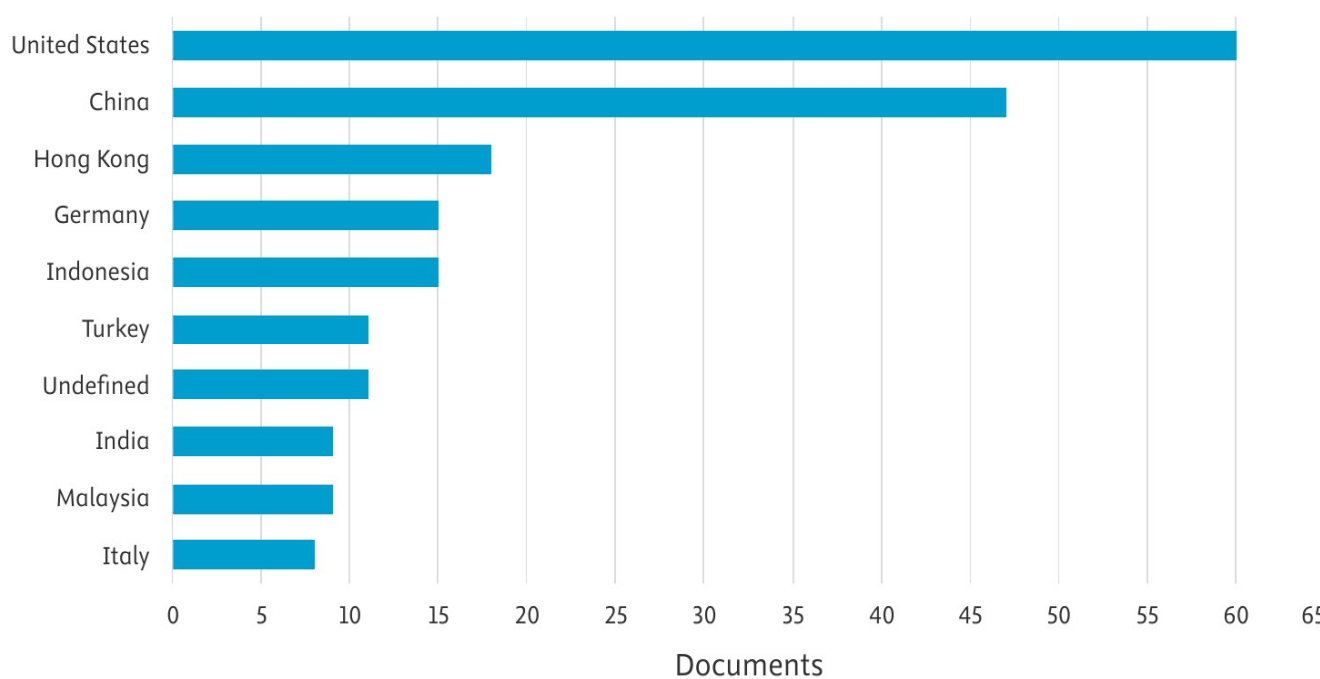


Figure 2 *Documents by Country or Territory*

Table 3: *Top Ten Countries and Territories by Number of Documents*

Rank	Country or territory	Documents	% of 294
1	United States	60	20.4
2	China	47	16.0
3	Hong Kong	18	6.1
4	Germany	15	5.1
4	Indonesia	15	5.1
6	Turkey	11	3.7
6	Undefined	11	3.7
8	India	9	3.1
8	Malaysia	9	3.1
10	Italy	8	2.7

RQ3: Productive Institutions and Authors

Institutions

Figure 3 lists the ten most productive affiliations. The Education University of Hong Kong was first with 9 documents, followed by Carnegie Mellon University (8), the University of Florida (7), the University of Utah (7), the University of Hong Kong (6), Carnegie Learning, Inc. (6), Central China Normal University (5), and Worcester Polytechnic Institute (5). Peking University and Khon Kaen University each had 4 documents.

The leading institution accounts for only 3.1% of all documents, so no single institution dominates this field. The list also shows a mix of university groups and one company (Carnegie Learning, Inc.), which develops mathematics learning software. Five of the ten affiliations are in the United States, two are in Hong Kong, two are in mainland China, and one is in Thailand.

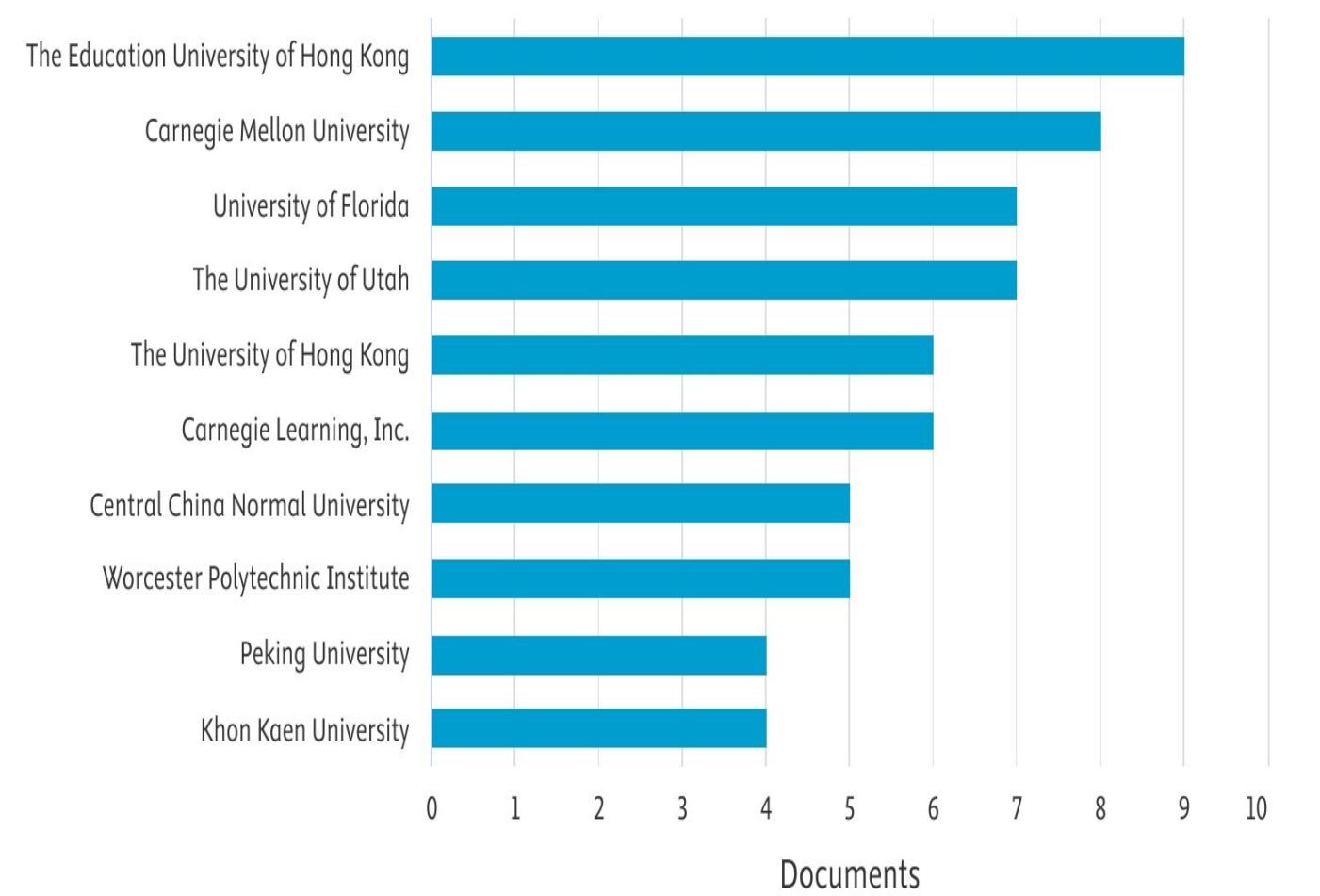


Figure 3: Documents by Affiliation

Authors

Figure 4 shows the most productive authors. Li, C. was first with 7 documents, followed by Xing, W. (6 documents). Fancsali, S. E., Li, H., Ritter, S., and Sun, D. each had 5 documents, and Aleven, V., Belpaeme, T., and Borchers, C. each had 4.

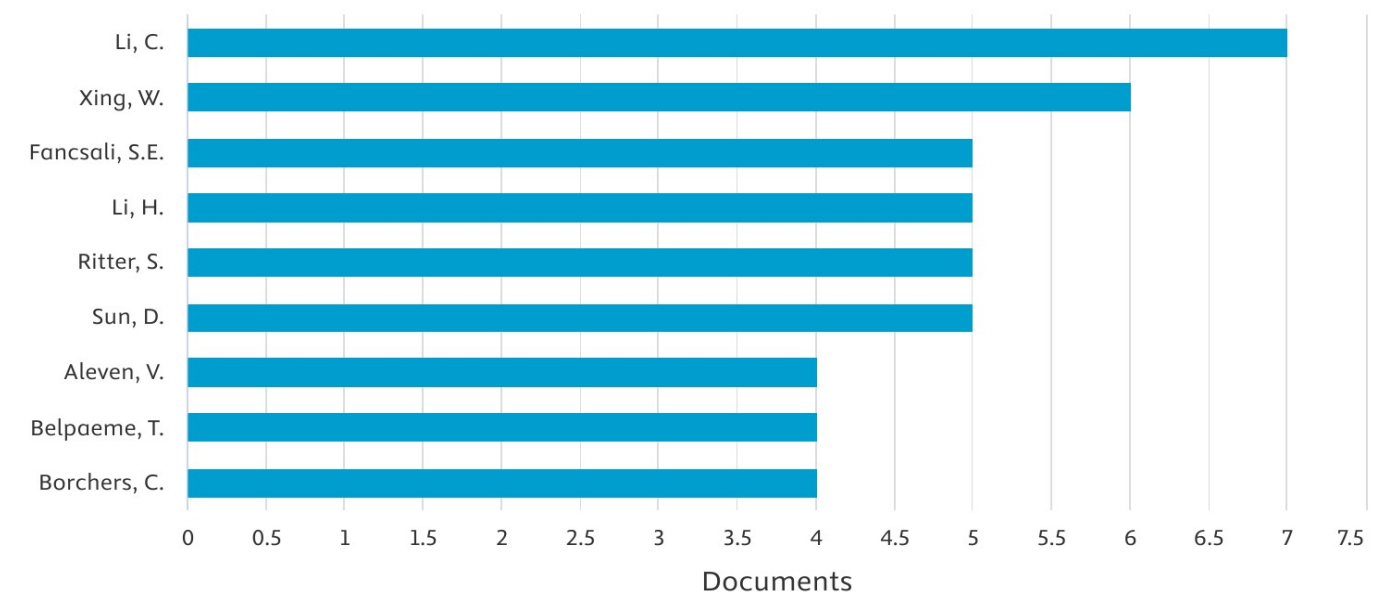


Figure 4: Documents by Author

Table 4: Main Research Groups behind the Most Productive Authors

Authors	Main affiliation	Main focus in the dataset	Example
Xing, W.; Li, C.; Li, H.	University of Florida (Xing); University of Utah (Li, C.)	Generative AI teachable agents and language models for middle school mathematics	Xing et al. (2025); Zhang et al. (2025)
Fancsali, S. E.; Ritter, S.	Carnegie Learning, Inc.	Embedded formative assessment and prediction of state test scores (2018–2020)	Zheng et al. (2019)
Aleven, V.; Borchers, C.	Carnegie Mellon University	Tutoring systems, caregiver support, and combining language models with tutoring systems	Venugopalan et al. (2025)
Belpaeme, T.	Ghent University	Language-model-powered social robots as math tutors	Rutatola et al. (2025)
Sun, D.	The Education University of Hong Kong	Generative AI learning environments for word problem solving	Liu et al. (2025)

The most productive author wrote only 7 of the 294 documents (2.4%). This low concentration means that the field is spread across many research teams. The titles of the documents in the Scopus database show that the top authors form five main groups, and each group works on a different part of the field (see Table 4). The

Carnegie Learning group (Fancsali and Ritter) published between 2018 and 2020, for example on the use of embedded formative assessment to predict state test scores (Zheng et al., 2019). In contrast, the groups around Xing and Li, around Sun, and around Belpaeme published in 2025 and 2026 and work with generative AI or large language models (Liu et al., 2025; Rutatola et al., 2025; Xing et al., 2025; Zhang et al., 2025).

RQ4: Funding Sponsors

Figure 5 shows the ten funding sponsors that appear most often. The National Science Foundation (NSF) of the United States supported 13 documents, and the National Natural Science Foundation of China (NSFC) supported 11. The Education University of Hong Kong, the Institute of Education Sciences, and the U.S. Department of Education each supported 5 documents. The National Research Council of Thailand and VLIRUOS (Belgium) each supported 4, and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil), the European Commission, and the Research Grants Council of Hong Kong each supported 3.

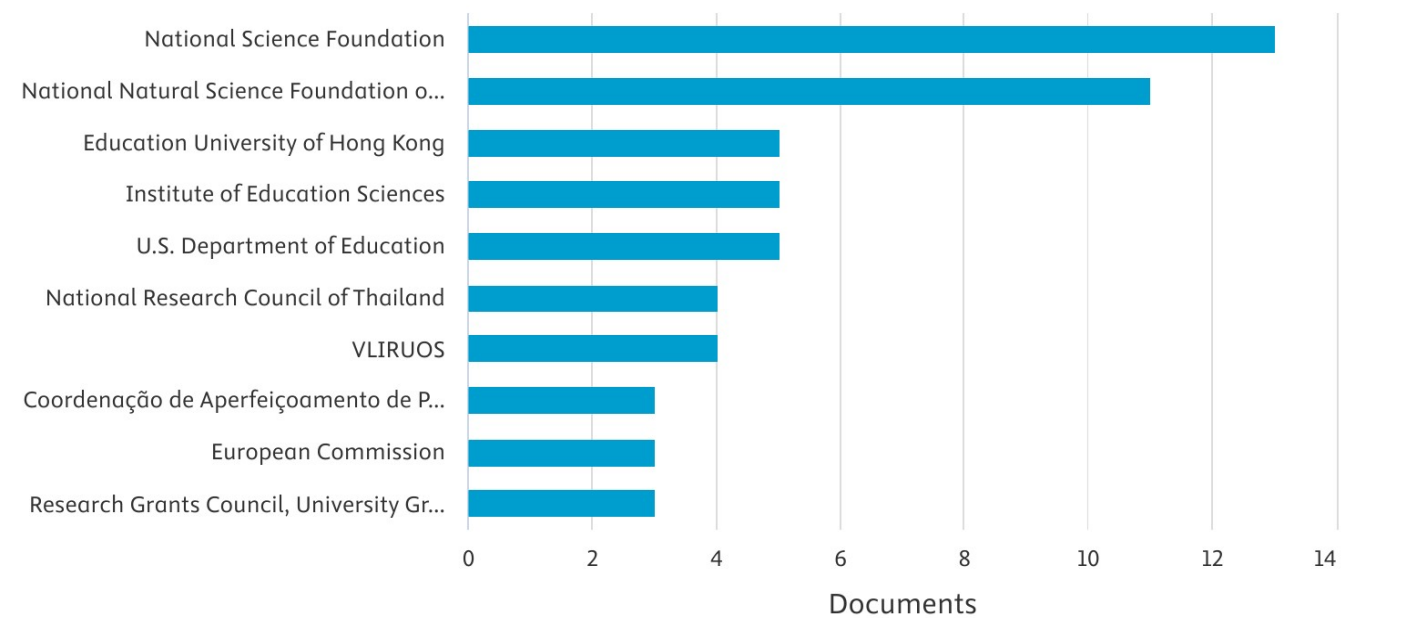


Figure 5: *Documents by Funding Sponsor*

The order of the top two funders matches the order of the top two countries, which suggests that national research funding is linked to research output in the United States and in China. Three United States federal bodies (NSF, the Institute of Education Sciences, and the Department of Education) appear 23 times in total, although some documents may be counted for more than one of them. The list also includes funders from Europe, South America, and Southeast Asia, which shows wide international interest. In addition, data shows that 137 of the 294 documents (46.6%) report some kind of funding, so more than half of the documents do not.

RQ5: Research Themes and Their Change over Time

The keyword analysis used the 250 documents that have author keywords. After joining spelling variants, “mathematics education” (82 documents) and “artificial intelligence” (78) were the most common keywords, which is expected because they are part of the search. Next came large language models (35), intelligent tutoring systems (26), generative AI (21), STEM or STEAM education (19), ChatGPT (15), and personalized learning (14) (see Table 5).

Figure 6 shows how the main themes changed. Terms about intelligent tutoring systems appear in 58 documents and are present in most years from 2016. Terms about ChatGPT, generative AI, and large language models are absent until 2023: only one document mentions them in 2023, but 19 do so in 2024, 66 in 2025, and 51 in 2026. In total, 137 documents (46.6%) mention at least one of these terms, and they make up about 57% of the documents in 2025. At the same time, 47 of the 58 tutoring-system documents do not mention generative AI, so classic tutoring research continues next to the new work.

The citation data also show the older base of the field. The most cited document is a study that compared an adaptive learning system with teacher-led instruction (Wang et al., 2023; 253 citations), followed by a study of AI education and computational thinking in a STEAM high school (Huang & Qiao, 2024; 113 citations) and a study of educational technology use during the COVID-19 pandemic (Christopoulos & Sprangers, 2021; 99 citations). Among the most cited documents are also an early study of tutoring systems as a math gap reducer in an after-school program (Huang et al., 2016) and two studies on ChatGPT in mathematics teaching (Egara & Mosimege, 2024; Getenet, 2024). The 294 documents have 2,132 citations in total, but the median is only one citation per document, so a few documents receive most of the citations. Many recent documents have had little time to be cited.

Table 5: *Most Frequent Author Keywords After Joining Variants*

Keyword	Documents	% of 250
Mathematics education (including “math education”)	82	32.8
Artificial intelligence (including “AI”)	78	31.2
Large language models (including “LLMs”)	35	14.0
Intelligent tutoring systems	26	10.4
Generative AI (including “generative artificial intelligence”)	21	8.4
STEM or STEAM education	19	7.6
ChatGPT	15	6.0
Personalized learning	14	5.6

Note. 250 of the 294 documents have author keywords.

Discussion

The results give a clear answer to the research questions of this study. Research on AI in school mathematics education was small for most of the last decade and then grew very fast after 2023. This timing matches the arrival of ChatGPT and other generative tools, and the keyword analysis supports this link: generative AI terms are almost absent before 2024 and then appear in about half of all documents. The field has therefore

moved from a period of intelligent tutoring systems to a period in which language models and generative AI dominate the discussion (Miao & Holmes, 2023).

The results also show who drives the field. The United States and China lead by a large margin, and their national funders, NSF and NSFC, are also the two largest sponsors. At the same time, the field is open and international. Hong Kong, Germany, Indonesia, Turkey, India, Malaysia, and Italy all appear among the top ten, and no single university or author is responsible for more than about 3% of the documents. Such a spread suggests that many teams are entering the field at the same time, which is usual for a topic in its early growth phase.

Implications

For researchers, the growth of generative AI studies means that new evidence will arrive quickly, and that reviews need to be updated often. There is space for more work from countries that are less visible in the top ten, and for studies that follow students over a longer time. For teachers and policy makers, the fast growth shows that the evidence base is still developing. Decisions about classroom use of AI tools should therefore use recent, careful studies and national guidance rather than early reports alone. For funders, the results show that most reported support comes from a small group of national agencies, and that more than half of the documents report no funding at all.

Limitations and Future Research

This study has limits. First, only Scopus database was used, and other databases such as Web of Science or Google Scholar may give different counts. Second, only English documents were included and review articles were excluded, so the field may be larger than shown here. Third, the year 2026 is incomplete, and recent documents have had little time to receive citations. Finally, this study used counts and simple trend measures. Future work can add science mapping, such as co-citation or keyword network analysis (Donthu et al., 2021), and can code the documents by school level, mathematics topic, and study design.

Conclusion

This bibliometric study analyzed 294 Scopus documents on AI in school mathematics education from 2016 to 2026. The main finding is a sharp rise in research after 2023, with 84% of all documents published in the last three years. The United States and China lead in output and funding, but the topic is spread across many countries, institutions, and authors. The themes moved from intelligent tutoring systems to ChatGPT and large language models. These results give researchers, teachers, and policy makers a clear map of the field and show where new studies are most needed.

References

- Baas, J., Schotten, M., Plume, A., Côté, G., & Karimi, R. (2020). Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies*, 1(1), 377–386. https://doi.org/10.1162/qss_a_00019
- Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 1, Article 100002. <https://doi.org/10.1016/j.caeai.2020.100002>

- Christopoulos, A., & Sprangers, P. (2021). Integration of educational technology during the Covid-19 pandemic: An analysis of teacher and student receptions. *Cogent Education*, 8(1), Article 1964690. <https://doi.org/10.1080/2331186X.2021.1964690>
- 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>
- Egara, F. O., & Mosimege, M. (2024). Exploring the integration of artificial intelligence-based ChatGPT into mathematics instruction: Perceptions, challenges, and implications for educators. *Education Sciences*, 14(7), Article 742. <https://doi.org/10.3390/educsci14070742>
- Getenet, S. (2024). Pre-service teachers and ChatGPT in multistrategy problem-solving: Implications for mathematics teaching in primary schools. *International Electronic Journal of Mathematics Education*, 19(1), Article em0766. <https://doi.org/10.29333/iejme/14141>
- Huang, X., Craig, S. D., Xie, J., Graesser, A., & Hu, X. (2016). Intelligent tutoring systems work as a math gap reducer in 6th grade after-school program. *Learning and Individual Differences*, 47, 258–265. <https://doi.org/10.1016/j.lindif.2016.01.012>
- Huang, X., & Qiao, C. (2024). Enhancing computational thinking skills through artificial intelligence education at a STEAM high school. *Science & Education*, 33(2), 383–403. <https://doi.org/10.1007/s11191-022-00392-6>
- Hwang, G.-J., & Tu, Y.-F. (2021). Roles and research trends of artificial intelligence in mathematics education: A bibliometric mapping analysis and systematic review. *Mathematics*, 9(6), Article 584. <https://doi.org/10.3390/math9060584>
- Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 1, Article 100001. <https://doi.org/10.1016/j.caeai.2020.100001>
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78. <https://doi.org/10.3102/0034654315581420>
- Liu, J., Sun, D., Sun, J., Wang, J., & Yu, P. L. H. (2025). Designing a generative AI enabled learning environment for mathematics word problem solving in primary schools: Learning performance, attitudes and interaction. *Computers and Education: Artificial Intelligence*, 9, Article 100438. <https://doi.org/10.1016/j.caeai.2025.100438>
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918. <https://doi.org/10.1037/a0037123>
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://doi.org/10.54675/EWZM9535>
- Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of Cognitive Tutor Algebra I at scale. *Educational Evaluation and Policy Analysis*, 36(2), 127–144. <https://doi.org/10.3102/0162373713507480>

- Rutatola, E. P., Stroeken, K., & Belpaeme, T. (2025). “Habari, colleague!”: A qualitative exploration of the perceptions of primary school mathematics teachers in Tanzania regarding the use of social robots. *Applied Sciences*, 15(15), Article 8483. <https://doi.org/10.3390/app15158483>
- Venugopalan, D., Yan, Z., Borchers, C., Lin, J., & Aleven, V. (2025). Combining large language models with tutoring system intelligence: A case study in caregiver homework support. In *Proceedings of the 15th International Conference on Learning Analytics and Knowledge* (pp. 373–383). Association for Computing Machinery. <https://doi.org/10.1145/3706468.3706516>
- Wang, S., Christensen, C., Cui, W., Tong, R., Yarnall, L., Shear, L., & Feng, M. (2023). When adaptive learning is effective learning: Comparison of an adaptive learning system to teacher-led instruction. *Interactive Learning Environments*, 31(2), 793–803. <https://doi.org/10.1080/10494820.2020.1808794>
- Xing, W., Song, Y., Li, C., Liu, Z., Zhu, W., & Oh, H. (2025). Development of a generative AI-powered teachable agent for middle school mathematics learning: A design-based research study. *British Journal of Educational Technology*, 56(5), 2043–2077. <https://doi.org/10.1111/bjet.13586>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*, 2021, Article 8812542. <https://doi.org/10.1155/2021/8812542>
- Zhang, F., Li, C., Henkel, O., Xing, W., Baral, S., Heffernan, N., & Li, H. (2025). Math-LLMs: AI cyberinfrastructure with pre-trained transformers for math education. *International Journal of Artificial Intelligence in Education*, 35(2), 509–532. <https://doi.org/10.1007/s40593-024-00416-y>
- Zheng, G., Fancsali, S. E., Ritter, S., & Berman, S. R. (2019). Using instruction-embedded formative assessment to predict state summative test scores and achievement levels in mathematics. *Journal of Learning Analytics*, 6(2), 153–174. <https://doi.org/10.18608/jla.2019.62.11>
- Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429–472. <https://doi.org/10.1177/1094428114562629>