Revolutionizing education: An extensive analysis of large language models integration

Nguyen Van Viet[1], Luong Thi Minh Hue[2], Nguyen The Vinh[3], Mong Quoc Tuan[4], Bui Anh Tu[5], Le Anh Tu[6], Phan Thi Cuc[7] and Nguyen Huu Khanh[8]

[1,2,3,4,5,6,7]*Thai Nguyen University of Information and Communication Technology, Vietnam*
[8]*Thai Nguyen University, Vietnam*
Corresponding email: nvviet@ictu.edu.vn

A B S T R A C T

Large Language Models have garnered significant attention from companies, universities, and research groups in recent times, driven by the abundance of data available for their training. However, little evidence has been conducted in the field of education, leaving a huge gap that needs to be filled. Therefore, the purpose of this article is to provide an overview of the use of new areas of artificial intelligence in the field of education. We use the PRISMA method to analyze the relevant contents in detail to gather data, covering articles collected in the contemporary period between January 2019 and 2024. Results from 54 reviewed publications indicated that trends of utilizing LLMs in education have increased significantly since 2022 and arXiv preprint is the most common repository for declaring researchers' ideas. The application of LLMs can support the achievement of learning objectives, enhance the quality and accuracy of assessments, and contribute to improving the educational environment as well as the practical application of various subjects. Seven limitations are identified and discussed, opening several avenues for future research agenda.

INTRODUCTION

Large Language Models are a type of complex deep learning AI, built by training on huge amounts of data. The basic convolutional set is a set of neural networks that have an encoder and a decoder with self-focusing capabilities. Encoders and decoders extract meaning from a string of text and understand the relationships between words and phrases within it. LLMs are capable of unsupervised training, although the main explanation More precisely, the converter performs self-learning. Through this process, the transducer learns to understand grammar, language, and basic knowledge. This allows data scientists to use GPUs to train transformer-based LLMs, significantly reducing training time.

This article aims to analyze the overview of the use of large language models in learning, practical practice within the scope of education. Education plays an important role in the development and progress of human society(Gan et al., 2023). The use of large language models in education is an advanced approach that helps realize personalized learning goals, supports effective assessment, and enhances learning and practice experiences. Since universities already have harsh penalties in place to punish plagiarism in other ways, it is natural to extend them to LLMs (Milano et al., 2023).

To solve the above problems, research groups and scientists have raised and presented the use of LLM in education offers many potentials and opportunities, but also requires careful preparation to overcome the existing difficulties and challenges. This is shown through many general and applied research articles related to LLMs in the field of education such as: Author Enkelejda Kasneci and research team, 2023 analyze the discussion of large language models and point out their prominence in the field of education (Kasneci et al., 2023), "Paraphrasing with large language models"(Witteveen & Andrews, 2019); Jeon and Lee (2023) provided a discussion analysis of teachers' use of LLM and its importance (Jeon & Lee, 2023); LLM in all areas, further research is needed to apply(Nguyen* and Nguyen 2024) .

In 2024, The use of Large Language Models in the scope of education quality improvement, higher education, general education continues to expand and adapt, reflecting both advancements in AI capabilities and a growing recognition of the need for thoughtful integration in educational settings. Newer LLMs, such as GPT-4, Phi-3 meets law (Khanh* et al. 2024) and other proprietary models, are now even more accessible and versatile, providing significant support in areas like personalized learning pathways, adaptive feedback, and student engagement. For instance, these models can analyze students' progress in real-time, identify areas of struggle, and suggest tailored resources, fostering a more customized learning experience.

Therefore, this article aims to solve the above problem by analyzing bibliographies of scientific articles on the use of LLMs in the field of education:

Question 1: How does the number of articles published between 2019 and 2024 relate to the use of LLM in education?
Question 2: What keywords appear most frequently in the journal and conference database?
Question 3: What are the most important research topics in the use of LLMs in the field of education?
Question 4: What are the gaps and areas for future research?
Answering the above research questions will help agencies in the field of education, universities and education systems gain basic perspectives to approach LLMs in the field of higher education learn. New researchers can find new future research directions through research gaps.

OBJECTIVES OF THE STUDY

This article aims to comprehensively evaluate LLMs (Language Model Models) in education and higher education. By reviewing relevant literature, we intend to shed light on the utilization and research trends of LLMs in educational contexts, with a focus on identifying future research directions.

The specific objectives of this study include:

1. Providing an overview of how LLMs are used in education, including an examination of pertinent publications.
2. Analyzing the themes, keywords, and publication timelines of LLMs research in the field of education.
3. Identifying the necessary modes and operational processes of LLMs within educational settings.

Our ultimate goal is to equip researchers with a foundational understanding of data analysis related to LLMs in within the scope of education, enabling them to shape and guide future research efforts.

MATERIALS AND METHODS

This study applied bibliometric analysis to evaluate the selected articles. Quantitative bibliometrics is understood as a field of library science that focuses on the application of mathematical and statistical analysis methods to bibliographic research, as well as performing numerical analyses of books, articles, and other types of publications. Bibliometric analysis has become a popular tool for quantitative evaluation of scientific research, which not only helps to determine the importance and influence of each research work but also evaluates the development and trends of research fields. The applied methods include publication volume analysis, author contribution analysis, co-authorship network analysis and keyword analysis. Using software such as VOSviewer, Microsoft Excel to support, the study identified prominent publication trends, collaborative relationships between authors and major research topics.

The article employs the PRISMA methodology for its research approach (Haddaway et al. 2022). In 2009, this method was published, to help reviewers evaluate what other authors did, and what they studied and published (Page et al. 2021). It offers a structured framework for conducting systematic and transparent analyses of scientific literature, aiding researchers in assessing the credibility of studies prior to integrating them into their own research endeavors. There are numerous of review paper was conducted by using this methodology.

The PRISMA methodology is a widely used scientific research method in systematic reviews and qualitative data analysis. The standard steps and guidelines in the PRISMA methodology include defining the research question, searching for relevant studies, selecting appropriate studies for analysis, drawing conclusions and synthesizing the selected studies, assessing the quality of the studies, and drawing conclusions. These steps are useful for systematic reviews of systematic reviews, literature analysis, and documents.

Search sources

The study primarily utilized Web of Science and Google Scholar as its primary data collection tool. Web of Science is a comprehensive, multidisciplinary research platform that provides access to a vast array of academic and scholarly resources. Web of Science, originally developed by the Institute for Scientific Information (ISI) and now maintained by Clarivate Analytics, is primarily known for indexing scholarly journals, conference proceedings, and patents across a variety of disciplines, including the natural sciences, social sciences…. Google Scholar provides versatile search features, enabling access to digital articles and documents online. This versatility not only facilitates swift data collection but also ensures a broad range of data sources, thereby enhancing the comprehensiveness of the study's evaluation and synthesis of information gathered from various outlets on WoS and Google Scholar.

Search criteria

The author selects articles for general analysis from journal and conference databases. Article selection criteria: i) Search terms: at least one term related to Large Language Models in Education appears in the article title;

ii) The term "Large Language Models in Education". The time to collect article data is from October 2019 to October 2024 since the 5-year time range is very common for a review paper. As a result, there are 54 articles that meet the requirements for inclusion in the analysis.

Conditions for including articles in analysis

To be included in the final analysis and evaluation, articles must fulfill additional requirements outlined as follows:

- Time: Articles must have been published from 2019 to the present.
- Language: Articles must be written in English.
- Accessibility: Full-text access to the article is required.

Articles failing to meet any of the following criteria will be excluded from the list:

- The article is not written in English.
- Articles published prior to 2019.
- The article pertains to artificial intelligence but not to education.
- The document is not an article (e.g., book, thesis, poster, introductory page, etc.).

Fig. 1 Diagram showing the movement of information through the different stages of a systematic review rocess utilizing the PRISMA method. Initially, 1610 records were identified through searches in Web of Science and Google Scholar. Subsequently, 1345 duplicate articles, those not classified as articles, those not in English language, and those unrelated to educaiton were removed. A total of 275 records were screened. Articles excluded based on exclusion criteria with review topic, abstract (158 records). The screening process led to 117 reports being selected for detailed assessment. However, 32 reports were not retrieved (e.g., due to access issues or missing documents). Of the 85 reports assessed for eligibility, 31 were excluded. Ultimately, 54 papers were included in this study for evaluation and analysis.
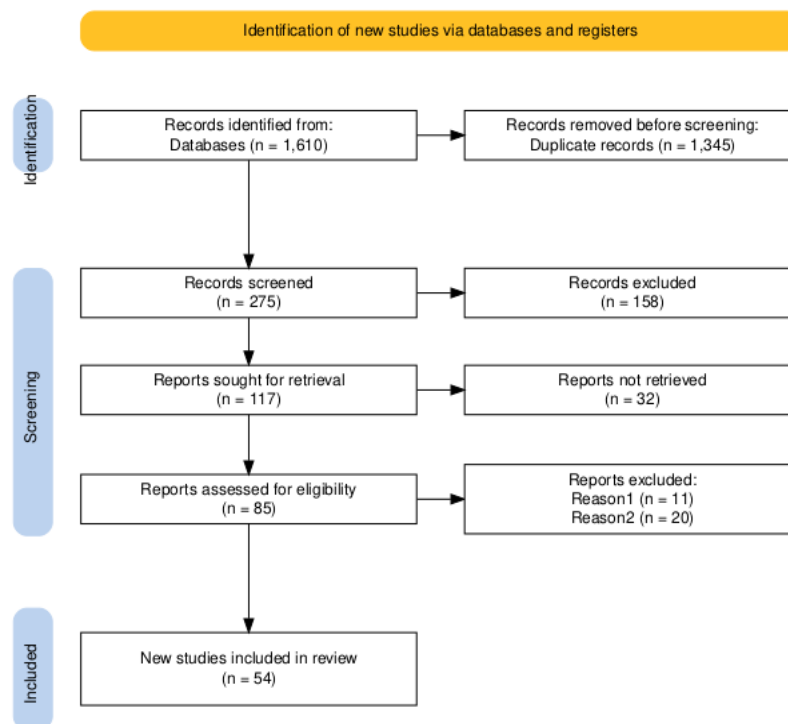

Fig. 1. Diagram showing the movement of information through the different stages of a systematic review

RESULTS AND DISCUSSION

How does the number of articles published between 2019 and 2024 relate to the use of LLM in education?

Fig. 2 presents an overview of publication trends on the use of Large Language Models (LLMs) in education from 2019 to 2024. A total of 54 articles were reviewed, showing a steady increase in publications over time. Between 2019 and 2022, the annual number of articles in this area was relatively low, with fewer than seven articles each year. However, 2023 marked a significant rise, with 29 articles associated with relevant keywords. This upward trend continued into 2024, with 19 articles identified.

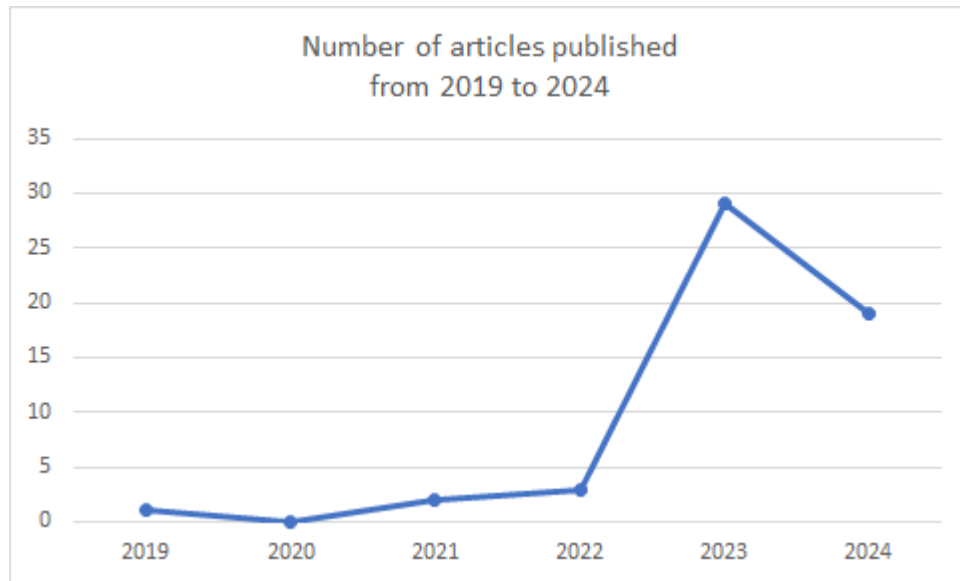Trends in LLM-related search terms from 2019 to 2024 include:

Fig. 2. Number of articles and publications published by year from 2019 to 2024

Table 1. Data table of articles published in the field of analytical articles

Journal source	Number of articles
Web of Science(Alshaikh, Al-Malki, and Almasre 2024; Benitez et al. 2024; Claman and Sezgin 2024; Diab Idris, Feng, and Dyo 2024; Gao, Wang, and Wang 2024; Guettala et al. 2024; Hennekeuser et al. 2024; Huang, Guo, and Zhang 2024; Jafari 2024; Jost, Taneski, and Karakatic 2024; Kang et al. 2024; Kyrpa et al. 2024; Lang et al. 2024; Luo and Yang 2024; Pan and Ni 2024; Pelaez-Sanchez, Velarde-Camaqui, and Glasserman-Morales 2024; Rajashekar et al. 2024; Ray 2024; Sun et al. 2024)	19
arXiv preprint(Caines et al., 2023; Choi et al., 2023; Hicke et al., 2023; Krupp et al., 2023; C. Li et al., 2023; Q. Li et al., 2023; Nguyen et al., 2023; Savelka et al., 2023; Tamkin et al., 2021; Witteveen & Andrews, 2019; Yan et al., 2023; Zhao et al., 2023)	12
International Conference on Artificial Intelligence in Education(Moore et al., 2023; Wang et al., 2022)	2
IEEE(Gan et al., 2023)	1
Authorea Preprints(Hadi et al., 2023)	1
BioData Mining(Meyer et al., 2023)	1
Computer(Ahmad et al., 2023)	1

Didactica Danubiensis(Morosanu et al., 2023)	1
Education for Chemical Engineers(Tsai et al., 2023)	1
Educational Dimension(Hamaniuk, 2021)	1
Int. J. Mod. Educ. Comput. Sci(Dao & Le, 2023)	1
JMIR Medical Education(Safranek et al., 2023)	1
Learning and individual differences(Kasneci et al., 2023)	1
Nature Machine Intelligence(Milano et al., 2023)	1
Nature Reviews Psychology(Demszky et al., 2023)	
Online Submission(Morrison, 2022)	1
BEA workshop (Xiao et al., 2023)	1
TESOL Quarterly(Pack & Maloney, 2023)	1
AJOB(Rahimzadeh et al., 2023)	1
Other(Agostini & Picasso, 2022; Baierl, 2023, Jeon & Lee, 2023, Bernabei et al., 2023...)	6

From Table 1 above, we see that Web of Science contributes the most articles related to Large language Models, Education with 19 articles accounting for 35% of analyzed articles. Ranked 2nd is the arXiv preprint with 12 articles accounting for 22% of analyzed articles. The remaining journals and conferences are nearly equivalent. This result provides useful information for researchers on LLMs in the fields of education, teaching, AI, ChatGPT... Publication rate on the Web of Science 35% in 2024 demonstrates rapid development and publication bringing artificial intelligence, large language models into the education field. Publication rate on the Arxiv platform, a pre-printed electronic database publications, without going through the peer review process, accounting for 22% can be explained by three main reasons. Firstly, these publications are all very new and have not been accepted for publication by reputable journals. Second, the authors want to receive comments from the community before officially publishing. Finally, the Declaration of sovereignty over the research field of the authors and author groups.

From Fig. 3, we can see that the research article on arXiv preprint by author Mark Chen and his colleagues is the most interesting and cited by scientists with a number of citations of 1064 in the googleScholar database (Kasneci et al., 2023). This shows the position and importance of their work in this field. Other articles researching LLMs in the field of education also receive a large number of citations such as by author Guettala, Pelaez-Sanchez, Velarde-Camaqui, and Glasserman-Morales, Liu; ... The remaining authors' articles have a citation count of 0 or 1, showing that not many people are interested in the author's articles.
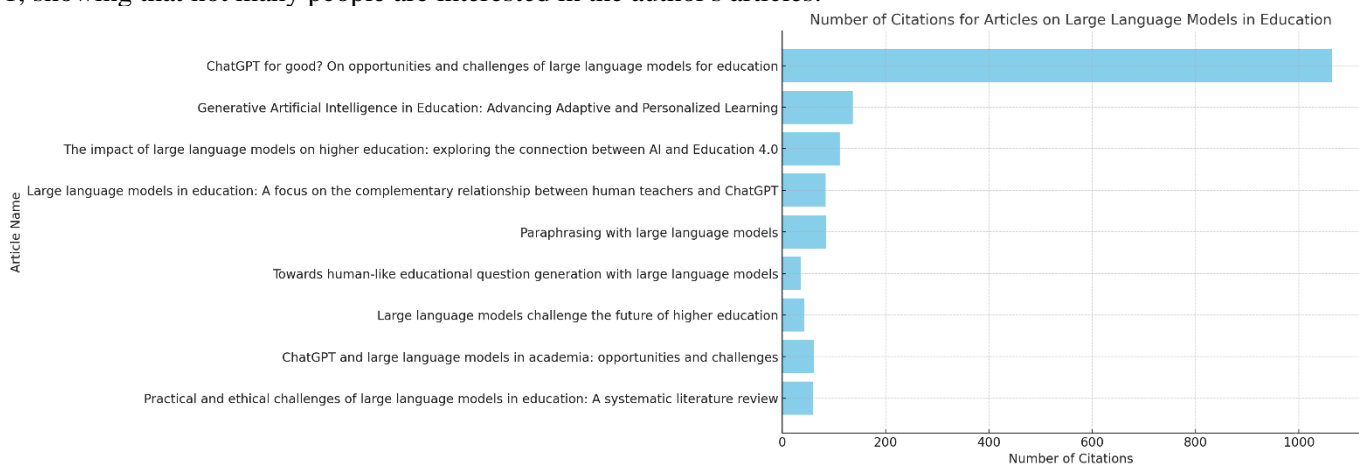

Fig. 3. Top articles with the most number of citations in the database related.

What keywords appear most frequently in the journal and conference database?

Fig. 4 shows a map visualizing research trends using Vosviewer software on the use of Large Language Models in the education sector from 2019 to 2024. The mapping results show important clusters focus on research topics on the impact of using LLMs on higher education, education....
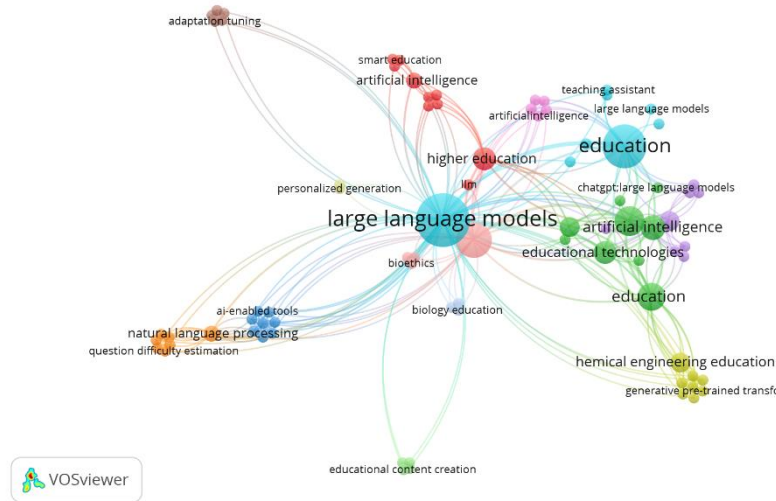

Fig. 4. Research areas related with keyword (LLMs, Education)

Fig. 5 shows the yearly distribution of content related to the Large Language Models research field. The purple color represents research conducted in 2022 in the fields of education, AI, and higher education . Yellow represents research conducted in the years from 2019 to 2024 in fields such as AI, Education technology, chemical engineering education...
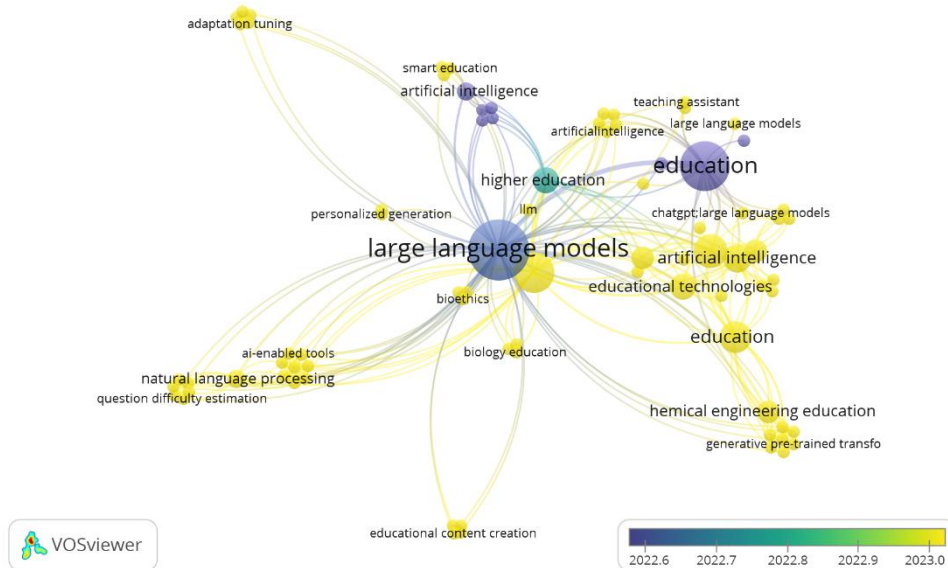

Fig. 5. Distribution by year of research fields related to Large Language Models

Fig. 6 and Fig. 7 show the most frequently occurring keywords such as Large Language Models with a total of 118 times; Education with 80 times ; ChatGPT with 49 times;.…

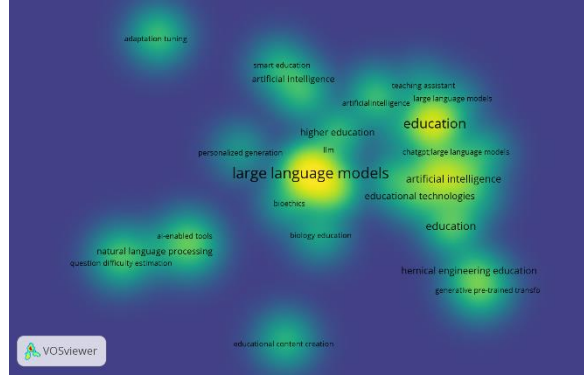*Fig. 6. Keywords were extracted from the abstracts of articles*

Fig. 7. The keyword cloud was extracted from the abstracts of articles

What are the most important research topics in the use of LLMs in the field of education?

Fig. 8 presents a synthesis of keywords used in the studies to help readers better understand the relationships between them. Based on the analysis table, clusters focus on content related to Large Language Models in the fields of education and higher education, as follows: Cluster 1 with 10 items, Cluster 2 with 9 items, and Cluster 3 with 8 items, among others. The keyword 'Large Language Models' has the highest frequency, appearing 118 times, followed by 'Education,' which ranks second with 80 occurrences...
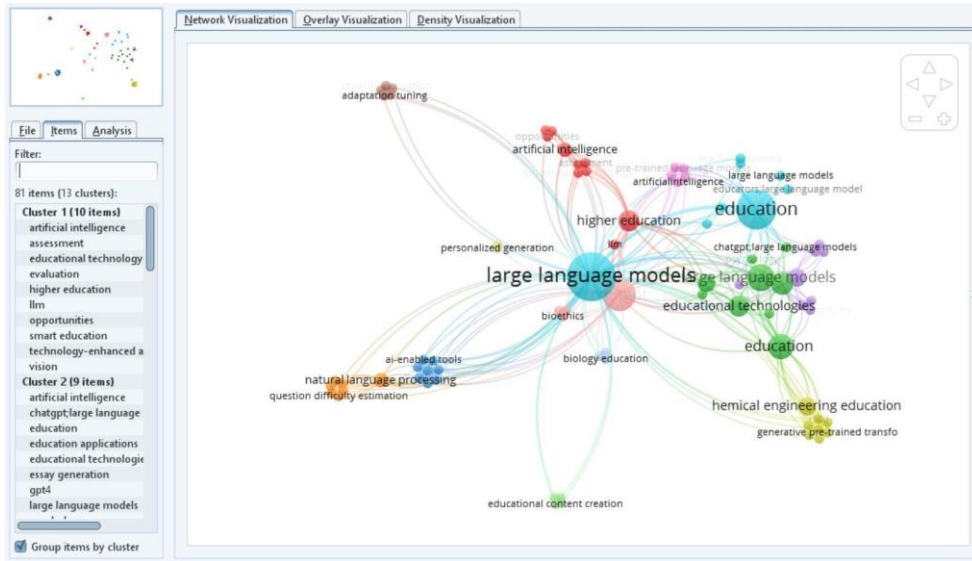
Fig. 8. Keyword occurrence analysis

These topics can be important future research directions to optimize the power of Large Language Models in the field of education.

What are the gaps and areas for future research?

Large language models in the field of education have been researched at a moderate level from 2019 to 2024 but by 2023 have gradually developed strongly and 2024 have 19 articles, but there are still some limitations and challenges that need to be overcome. Table 2 shows the frequency of limitations and challenges of Large language models in the education sector as identified in various publications.

Table 2. Frequency of limitations indicated in publications

Limitations/challenges	Reference publications
Safety issues	(Tamkin et al., 2021)
Honest review	(Morrison, 2022)
Difficult quality assessment, program evaluation	(Morrison, 2022; Savelka et al., 2023)
Misleading content	(Tamkin et al., 2021)
Structure generation	(Witteveen & Andrews, 2019)
Refine domain knowledge data only on source models	(Nguyen et al., 2023)
Limitations of data.	(Nguyen et al., 2023)

CONCLUSION AND RECOMMENDATION

In this article, research on Large Language Models in the fields of education and higher education with the purpose of analyzing and evaluating articles and documents related to the use of LLMs from 2019 to 2024. Research results show that the period from 2019 to 2022 has a low number of research on LLMs at 11% , skyrockets in 2023 with 54% and in 2024 with 35% in the fields of education, higher education, ChatGPT, …The top publications are "arXiv preprint", followed by " International Conference on Artificial Intelligence in Education " in researching LLMs in the field of education and the article "ChatGPT for good? On opportunities and challenges of large language models for education" is the most influential 1064 times cited LLMs study in the field of education.
The use of Large Language Models (LLMs) in education and higher education has significantly reshaped approaches to teaching, learning, and academic management. LLMs show great potential to improve educational outcomes by offering personalized learning, automated feedback, virtual tutoring, and support for administrative tasks. By providing adaptive resources, they help address individual learning needs and offer scalable support to both students and educators.

However, some challenges persist. Issues like potential biases, ethical concerns, and effects on students' critical thinking highlight the need for responsible, carefully monitored application. Over-reliance on AI tools raises the importance of a balanced approach, where LLMs enhance education but do not replace essential human interaction and oversight. Studies on frameworks for validating content generated by LLMs may help increase accuracy and trustworthiness.

This review provides a foundation for identifying gaps in current research and offers a direction for future studies to advance LLM applications in the educational field.

REFERENCES

Agostini, D., & Picasso, F. (2022). *Large Language Models for Sustainable Assessment and Feedback in Higher Education: Towards a Pedagogical and Technological Framework*.

Ahmad, N., Murugesan, S., & Kshetri, N. (2023). Generative Artificial Intelligence and the Education Sector. *Computer*, *56*(6), 72–76.

Alshaikh, R., Al-Malki, N., & Almasre, M. (2024). The implementation of the cognitive theory of multimedia learning in the design and evaluation of an AI educational video assistant utilizing language models. *Heliyon, 10*(3). https://doi.org/10.1016/j.heliyon.2024.e25361

Baierl, J. D. (2023). *Applications of Large Language Models in Education: Literature Review and Case Study*. University of California, Los Angeles.

Benitez, T. M., Xu, Y., Boudreau, J. D., Kow, A. W. C., Bello, F., Le Van Phuoc, Wang, X., & Sun, X., et al. (2024). Harnessing the potential of large language models in medical education: Promise and pitfalls. *Journal of the American Medical Informatics Association, 31*(3), 776–783. https://doi.org/10.1093/jamia/ocad252

Bernabei, M., Colabianchi, S., Falegnami, A., & Costantino, F. (2023). Students' use of large language models in engineering education: A case study on technology acceptance, perceptions, efficacy, and detection chances. *Computers and Education: Artificial Intelligence*, *5*, 100172.

Caines, A., Benedetto, L., Taslimipoor, S., Davis, C., Gao, Y., Andersen, O., Yuan, Z., Elliott, M., Moore, R., Bryant, C., & others. (2023). On the application of large language models for language teaching and assessment technology. *ArXiv Preprint ArXiv:2307.08393*.

Chhina, S., Antony, B., & Firmin, S. (2023). *Navigating the Terrain of Large Language Models in Higher Education-A systematic literature review*.

Choi, J. H., Garrod, O., Atherton, P., Joyce-Gibbons, A., Mason-Sesay, M., & Björkegren, D. (2023). Are LLMs Useful in the Poorest Schools? theTeacherAI in Sierra Leone. *ArXiv Preprint ArXiv:2310.02982*.

Claman, D., & Sezgin, E. (2024). Artificial intelligence in dental education: Opportunities and challenges of large language models and multimodal foundation models. *JMIR Medical Education, 10*. https://doi.org/10.2196/52346

Dao, X.-Q., & Le, N.-B. (2023). LLMs Performance on Vietnamese High School Biology Examination. *Int. J. Mod. Educ. Comput. Sci*, *15*(6), 14–30.

Demszky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., Eichstaedt, J. C., Hecht, C., Jamieson, J., Johnson, M., & others. (2023). Using large language models in psychology. *Nature Reviews Psychology*, *2*(11), 688–701.

Diab Idris, M., Feng, X., & Dyo, V. (2024). Revolutionizing higher education: Unleashing the potential of large models for strategic transformation. *IEEE Access, 12*, 67738–67757. https://doi.org/10.1109/ACCESS.2024.3400164

Gan, W., Qi, Z., Wu, J., & Lin, J. C.-W. (2023). Large language models in education: Vision and opportunities. *2023 IEEE International Conference on Big Data (BigData)*, 4776–4785.

Gao, Y., Wang, Q., & Wang, X. (2024). Exploring EFL university teachers' beliefs in integrating ChatGPT and other large language models in language education: A study in China. *Asia Pacific Journal of Education, 44*(1, SI), 29–44. https://doi.org/10.1080/02188791.2024.2305173

Guettala, M., Bourekkache, S., Kazar, O., & Harous, S. (2024). Generative artificial intelligence in education: Advancing adaptive and personalized learning. *Acta Informatica Pragensia, 13*(3), 460–489. https://doi.org/10.18267/j.aip.235

Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J., Mirjalili, S., & others. (2023). A survey on large language models: Applications, challenges, limitations, and practical usage. *Authorea Preprints*.

Hamaniuk, V. A. (2021). The potential of Large Language Models in language education. *Educational Dimension*, *5*, 208–210.

He, J., & Vechev, M. (2023). Large language models for code: Security hardening and adversarial testing. *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security*, 1865–1879.

Hennekeuser, D., Vaziri, D. D., Golchinfar, D., Schreiber, D., & Stevens, G. (2024). Enlarged education: Exploring the use of generative AI to support in higher education. *International Journal of Artificial Intelligence in Education.* https://doi.org/10.1007/s40593-024-00424-y

Hicke, Y., Agarwal, A., Ma, Q., & Denny, P. (2023). ChaTA: Towards an Intelligent Question-Answer Teaching Assistant using Open-Source LLMs. *ArXiv Preprint ArXiv:2311.02775*.

Huang, S., Guo, M., & Zhang, S. (2024). Research on the construction and optimization of the educational system of intelligent educational technology in the context of knowledge economy. *Journal of the Knowledge Economy.* https://doi.org/10.1007/s13132-024-01767-6

Jafari, E. (2024). Artificial intelligence and learning environment: Human considerations. *Journal of Computer Assisted Learning.* https://doi.org/10.1111/jcal.13011

Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 1–20.

Jost, G., Taneski, V., & Karakatic, S. (2024). The impact of large language models on programming education and student outcomes. *Applied Sciences (Basel), 14*(10). https://doi.org/10.3390/app14104115

Kang, K., Yang, Y., Wu, Y., & Luo, R. (2024). Integrating large language models in bioinformatics education for medical students: Opportunities and challenges. *Annals of Biomedical Engineering, 52*(9), 2311–2315. https://doi.org/10.1007/s10439-024-03554-5

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., & others. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274.

Khanh, N. H., Viet, N. V., Vinh, N. T., & Cong, N. H. (2024). Phi-3 meets law: Fine-tuning mini language models for legal document understanding. *Research, Development and Application on Information and Communication Technology*, *2024*(3), 136–142.

Krupp, L., Steinert, S., Kiefer-Emmanouilidis, M., Avila, K. E., Lukowicz, P., Kuhn, J., Küchemann, S., & Karolus, J. (2023). Challenges and Opportunities of Moderating Usage of Large Language Models in Education. *ArXiv Preprint ArXiv:2312.14969*.

Kyrpa, A., Stepanenko, O., Zinchenko, V., Datsiuk, T., Karpan, I., & Tilniak, N. (2024). Artificial intelligence tools in teaching social and humanitarian. *Information Technologies and Learning Tools, 100*(2), 162–179. https://doi.org/10.33407/itlt.v100i2.5563

Lang, Q., Tian, S., Wang, M., & Wang, J. (2024). Exploring the answering capability of large language models in addressing complex knowledge in entrepreneurship education. *IEEE Transactions on Learning Technologies, 17*, 2107–2116. https://doi.org/10.1109/TLT.2024.3456128

Li, C., Zhang, M., Mei, Q., Wang, Y., Hombaiah, S. A., Liang, Y., & Bendersky, M. (2023). Teach LLMs to Personalize– An Approach inspired by Writing Education. *ArXiv Preprint ArXiv:2308.07968*.

Li, Q., Fu, L., Zhang, W., Chen, X., Yu, J., Xia, W., Zhang, W., Tang, R., & Yu, Y. (2023). Adapting Large Language Models for Education: Foundational Capabilities, Potentials, and Challenges. *ArXiv Preprint ArXiv:2401.08664*.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., & Liu, Z. (2023). Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-Radiology*, 100017.

Luo, Y., & Yang, Y. (2024). Large language model and domain-specific model collaboration for smart. *Frontiers of Information Technology & Electronic Engineering, 25*(3), 333–341. https://doi.org/10.1631/FITEE.2300747

Meyer, J. G., Urbanowicz, R. J., Martin, P. C. N., O'Connor, K., Li, R., Peng, P.-C., Bright, T. J., Tatonetti, N., Won, K. J., Gonzalez-Hernandez, G., & others. (2023). ChatGPT and large language models in academia: opportunities and challenges. *BioData Mining*, *16*(1), 20.

Milano, S., McGrane, J. A., & Leonelli, S. (2023). Large language models challenge the future of higher education. *Nature Machine Intelligence*, *5*(4), 333–334.

Moore, S., Tong, R., Singh, A., Liu, Z., Hu, X., Lu, Y., Liang, J., Cao, C., Khosravi, H., Denny, P., & others. (2023). Empowering education with llms-the next-gen interface and content generation. *International Conference on Artificial Intelligence in Education*, 32–37.

Morosanu, G.-A., Rata, L.-A., Geru, M., & others. (2023). Outcomes of Large Language Models and Artificial Intelligence in Education. *Didactica Danubiensis*, *3*(1), 30–52.

Morrison, R. (2022). Large Language Models and Text Generators: An Overview for Educators. *Online Submission*.

Nguyen, V. V., & Nguyen, V. T. (2024). Large language models in software engineering: A systematic review and vision. *Journal of Engineering and Software Innovation, 2*(2). https://ejournal.papanda.org/index.php/jesi/article/view/968

Nguyen, M. H., Tschiatschek, S., & Singla, A. (2023). Large Language Models for In-Context Student Modeling: Synthesizing Student's Behavior in Visual Programming from One-Shot Observation. *ArXiv Preprint ArXiv:2310.10690*.

Pack, A., & Maloney, J. (2023). Using Generative Artificial Intelligence for Language Education Research: Insights from Using OpenAI's ChatGPT. *TESOL Quarterly*, *57*(4), 1571–1582.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, n71*. https://doi.org/10.1136/bmj.n71

Pan, G., & Ni, J. (2024). A cross-sectional investigation of ChatGPT-like large language models among medical students in China. *BMC Medical Education, 24*(1). https://doi.org/10.1186/s12909-024-05871-8

Pelaez-Sanchez, I. C., Velarde-Camaqui, D., & Glasserman-Morales, L. D. (2024). The impact of large language models on higher education: Exploring the connection between AI and Education 4.0. Frontiers in Education, 9. https://doi.org/10.3389/feduc.2024.1392091

Pfeffer, O. P., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *2023-01-15]. Https://"阿拉丁神灯" 还是 "潘多拉魔盒": ChatGPT 教育应用的潜能与风险现代远程教育研究*, *35*(2).

Rahimzadeh, V., Kostick-Quenet, K., Blumenthal Barby, J., & McGuire, A. L. (2023). Ethics education for healthcare professionals in the era of chatGPT and other large language models: Do we still need it? *The American Journal of Bioethics*, *23*(10), 17–27.

Rajashekar, N., Chan, C., Laine, L., & Shung, D. (2024). Large language model-based simulated patients with upper bleeding for medical education: A pilot study with. *Gastroenterology, 166*(5, S), S933.

Ray, P. P. (2024). Embracing the AI revolution with open large language models in anatomy. *Surgical and Radiologic Anatomy, 46*(7), 949–950. https://doi.org/10.1007/s00276-024-03379-w

Safranek, C. W., Sidamon-Eristoff, A. E., Gilson, A., & Chartash, D. (2023). The role of large language models in medical education: applications and implications. In *JMIR Medical Education* (Vol. 9, p. e50945). JMIR Publications Toronto, Canada.

Savelka, J., Agarwal, A., Bogart, C., & Sakr, M. (2023). From GPT-3 to GPT-4: On the Evolving Efficacy of LLMs to Answer Multiple-choice Questions for Programming Classes in Higher Education. *ArXiv Preprint ArXiv:2311.09518*.

Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., Scales, N., Tanwani, A., Cole-Lewis, H., & Pfohl, S. (2023). Large language models encode clinical knowledge. *Nature*, *620*(7972), 172–180.

Sun, J., Wei, M., Feng, J., Yu, F., Li, Q., & Zou, R. (2024). Progressive knowledge tracing: Modeling learning process from abstract to concrete. *Expert Systems with Applications, 238*(F). https://doi.org/10.1016/j.eswa.2023.122280

Tamkin, A., Brundage, M., Clark, J., & Ganguli, D. (2021). Understanding the capabilities, limitations, and societal impact of large language models. *ArXiv Preprint ArXiv:2102.02503*.

Tsai, M.-L., Ong, C. W., & Chen, C.-L. (2023). Exploring the use of large language models (LLMs) in chemical engineering education: Building core course problem models with Chat-GPT. *Education for Chemical Engineers*, *44*, 71–95.

Wang, Z., Valdez, J., Basu Mallick, D., & Baraniuk, R. G. (2022). Towards human-like educational question generation with large language models. *International Conference on Artificial Intelligence in Education*, 153–166.

Witteveen, S., & Andrews, M. (2019). Paraphrasing with large language models. *ArXiv Preprint ArXiv:1911.09661*.

Xiao, C., Xu, S. X., Zhang, K., Wang, Y., & Xia, L. (2023). Evaluating reading comprehension exercises generated by LLMs: A showcase of ChatGPT in education applications. *Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, 610–625.

Xing, W., Lan, A., Crossley, S., Yu, Z., Denny, P., Nixon, N., & Stamper, J. (n.d.). *Call for Papers: Special Issue on the Use of Large Language Models in Education*.

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2023). Practical and ethical challenges of large language models in education: A systematic literature review. *ArXiv Preprint ArXiv:2303.13379*.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., & others. (2023). A survey of large language models. *ArXiv Preprint ArXiv:2303.18223*.