Scientometrica 1 (2025) 153–164

Contents lists available at GrowingScience

Scientometrica

homepage: www.GrowingScience.com/sci

A scientometrics survey of highly cited research on large language models in education

Reza Ghaeli[a*]

[a]*Instructor, Computing Studies and Information Systems (CSIS), Faculty of Commerce & Business Administration (CBA), Douglas College, New Westminster, BC, Canada*

C H R O N I C L E

Article history:
Received: November 5, 2025
Received in revised format: November 5, 2025
Accepted: November 31, 2025
Available online:
November 31, 2025

Keywords:
Large Language Models
ChatGPT
Generative AI
Education
Scientometrics
Systematic Review
Artificial Intelligence in Education
Research Trends

A B S T R A C T

The emergence of large language models (LLMs), like ChatGPT, has caused a major shift in all academic and professional fields; education being the one impacted the most deeply. The first 200 most cited articles from Scopus on the topic of "large language models" and "education" were selected for a scientometric analysis of this survey. The aim is to map the intellectual landscape, specifying dominant research themes, key contributors, methodological trends, and the emergent challenges that researchers faced during this initial, explosive phase of research. This review systematically analyzes the publication trends, authorship patterns, geographical and institutional contributions, and the conceptual structure of the literature to provide a quantitative and qualitative snapshot of a field that is rapidly changing. The analysis uncovers a domain where exploratory studies, conceptual debates, and early empirical validations dominate, and where the focus of research is mainly on higher education and medical training. The major issues being discussed are academic integrity, assessment redesign, and the ethical integration of AI, while promising applications are personalized learning, automated feedback, and content generation. This survey is a reference for a better understanding of the roots and forefronts of LLM-related educational research as it passes from the initial reaction to systematic integration.

© 2025 by the authors; licensee Growing Science, Canada.

1. Introduction

The advanced large language models (LLMs), among them OpenAI's ChatGPT, which went public with heavy technology in late 2022, have come to the forefront as a major milestone in the course of developing artificial intelligence. The models extract their human-like text production and comprehension skills from being trained on huge text databases (Brown et al., 2020). The education sector, being oral and written communications and knowledge sharing determine, was clearly the first candidate for both interruption and transformation. The examples of LLMs as extremely personalized tutors, research assistants, content writers, and testing tools have raised as well fears about academic honesty, the decline of conventional assessment methods, and the spreading of biases and inaccuracies. The mentioned period covering late 2022 to 2025 has been a lively tempo of scholarly research. Researchers and educators from all over the world have been in a hurry to record, analyze and criticize the capabilities, implications, and potential uses of LLMs in the education sector. This has led to an ever-growing literature on the subject, as can be seen from the high citation counts of early works that are indicative of their major impact on the discourse.

This article introduces a scientometric analysis of this rapidly growing area. Rather than a conventional systematic review, which collects and presents evidence on a particular research question, a scientometric study quantitatively analyzes the literature and exposes its structural and dynamic features (Ellegaard & Wallin, 2015). This survey, through the examination of a curated dataset of 200 highly cited articles, intends to:

1. Quantify the growth and distribution of publications.

* Corresponding author.
E-mail address: ghaelim@douglascollege.ca (R. Ghaeli)

ISSN 3115-8455 (Online) - ISSN 3115-8447 (Print)
© 2025 by the authors; licensee Growing Science, Canada
doi: 10.5267/j.sci.2025.5.002

2. Determine the most influential authors, institutions, countries, and journals.
3. Map the key study themes and topical clusters.
4. Investigate the methodological methods and application domains.
5. Synthesize the predominant benefits and challenges determined in the literature.

The analysis results will reveal the academic community's first response to the LLM revolution in education and will show the areas of agreement, research gaps, and future directions.

2. Methodology

2.1 Data Collection and Source

The survey's data comes from a BibTeX file containing the bibliographic records of the 200 most highly cited articles. The primary search was conducted in the Scopus database using the keywords "large language models" AND "education." The results were ordered according to the citation count and the top 200 entries were gathered, indicating the most impactful publications at the time of the search. The dataset consists of articles published in 2022-2025 period, but most of them are in 2023-2024.

2.2 Data Analysis

In order to respond to the research question regarding the impact of AI technologies on the international division of production, there is going to be two qualitative methods of seeking answers to the research questions.

- **Quantitative Analysis:** The bibliographic data was analyzed to extract and count key metrics, including:
    - **Publication Year:** To track the temporal evolution of the field.
    - **Source Journal/Conference:** To determine the primary outlets for this research.
    - **Authors and Affiliations:** To determine prolific contributors and their institutions.
    - **Document Type:** To classify the nature of the publications (e.g., article, review, conference paper).
    - **Keywords:** To determine the most frequently discussed topics.
- **Qualitative Analysis:** The titles, abstracts, and author keywords of the 200 articles were analyzed in a review process to extract the main research themes, the respective application domains and the arguments about opportunities and challenges that made the most frequent appearances. This was an iterative reading, coding and categorization process that resulted in the construction of a clear thematic framework.

2.3 Limitations

The entire survey draws from one, though very strong, database (Scopus) and the most cited 200 articles. In this way, the highest-impact literature is concentrated, but at the same time, it might overlook the very new, specialized, or local research studies that have not received many citations yet. Moreover, the research area is changing very fast, so this study only shows the situation at one particular moment, and more trends will undoubtedly keep coming up.

3. Results and Analysis

3.1 Publication Trends and Growth

The scrutiny of publication years among the dataset-based data clearly demonstrates the rapid rise in the demand for LLMs in the educational sector right after the introduction of ChatGPT. The figures, which are compactly presented in Table 1, reveal a notable increase from just one pioneering conference paper in 2022 to the crest of publishing in 2023 and 2024. The two 2022 publications (MacNeil et al., 2022; Sarsa et al., 2022) were ahead of their time, investigating GPT-3 applications for generating code explanations and programming exercises, which in turn, helped to create the large amount of research that came along. The year 2023 marked the "reaction phase" where the academic world raised up the implications of ChatGPT and thus a great number of exploratory studies, commentaries, and early empirical tests followed. The large

number of publications in 2024 points to the maturing of the dialogue, with more organized empirical studies, systematic reviews, and inquiries into particular applications and ethical frameworks. The fact that there are 11 articles from 2025 already receiving huge citations, shows that this trend is going to continue; with research getting more specialized and focused on the aspects of integration and long-term impact.

Table 1
Distribution of Publications by Year

Year	Number of Publications	Percentage of Total
2022	2	1.00%
2023	97	48.50%
2024	90	45.00%
2025	11	5.50%
Total	**200**	**100.00%**

3.2 Document Types and Research Focus

The kinds of publications indicate the maturity level of the research area. In Table 2, we can see that journal articles are the major part of the literature, yet a large part is made up of reviews and conference papers, which are characteristics of quick, multi-disciplinary fields.

Table 2
Distribution of Document Types

Document Type	Number	Percentage
Article	134	67.00%
Review	31	15.50%
Conference Paper	26	13.00%
Note/Editorial/Letter	9	4.50%
Total	**200**	**100.00%**

Astonishingly, the extremely high proportion of Reviews (15.5%) speaks even louder. It shows that the whole research community is making a united effort to organize and understand the mountain of evidence that is being amassed quickly. Thus, the ground-breaking reviews like those of Yan et al. (2024), Sallam (2023), and Bozkurt (2023) have greatly contributed to the structuring of the dialogue. The large amount of Conference Papers (13.0%), mostly coming from computer education conferences (such as SIGCSE, ICER, ITiCSE), indicates the field's origin in computer science and the rapidity of sharing the findings. The Notes and Editorials are typically the publication of early and critical standpoints of journal editors and opinion leaders, which helps to define the future research agenda (Eysenbach, 2023; Kasneci et al., 2023; Meyer et al., 2023).

A qualitative analysis of the research focus shows a clear progression:

- **Initial Capability Assessment (Late 2022 - 2023):** A big bunch of initial research concentrated on the evaluation of ChatGPT and other large language models through standardized tests mainly in healthcare (like USMLE, medical board exams) (Gilson et al., 2023; Kung et al., 2023) and other professional areas (programming, engineering).

- **Implication and Integration Discussions (2023 - 2024):** At the same time as the benchmarking studies, the literature that covered widest areas and subjects such as opportunities, challenges and ethical implications of using AI in education (Dwivedi et al., 2023; Kasneci et al., 2023) was built up.

- **Applied Research and Pedagogical Innovation (2024 - 2025):** Recent activities have focused on creating, employing, and assessing particular LLM-based tools and teaching methods, including AI-driven feedback systems (Meyer et al., 2024), personalized learning aides (Sajja et al., 2024), and innovative evaluation approaches such as "prompt problems" (Denny et al., 2024).

3.3 Leading Journals and Conferences

The research dissemination on LLMs in education is characterized by a variety of venues, which demonstrate its interdisciplinary nature in a very broad manner. The major sources in the dataset are enumerated in Table 3. The fact that JMIR Medical Education takes the leading position points to the enormous effect and instant usage of LLMs in medical education and healthcare, which is going to be further discussed in the following sections. The presence of well-known educational technology journals (Education and Information Technologies, Computers and Education) and specialized AIEd journals shows that the issue of LLMs in the educational field is not only being discussed among the core educational researchers but is also gaining importance. The number of publications in globally reputed multidisciplinary journals such as Nature and Scientific Reports is high, which indicates that the technological shift is of great importance to society at large. The

strong participation in conferences related to computer science education indicates that programming education has been the main depot for testing LLM applications.

Table 3
Top Publication Venues

Venue Name	Number of Publications	Primary Discipline
JMIR Medical Education	12	Medical Education / Health Informatics
Education and Information Technologies	10	Educational Technology
Computers and Education: Artificial Intelligence	8	AI in Education
Computers and Education	5	Educational Technology
International Journal of Artificial Intelligence in Education	4	AI in Education
British Journal of Educational Technology	4	Educational Technology
Journal of Medical Systems	4	Medical Informatics
Scientific Reports	4	Multidisciplinary Science
Nature and its sister journals (*Communications Medicine*)	4	Multidisciplinary Science
ACM SIGCSE, ICER, ITiCSE Proceedings	12 (combined)	Computer Science Education

3.4 Influential Authors and Collaborative Networks

The data indicates a research environment that is very collaborative. Although there are only a handful of authors who publish quite often, the field is not exception to one single group only but it is characterized by various prolific contributors and new networks coming up. The creators of the publications in the dataset are mentioned in Table 4 if they have three or more publications.

Table 4
Most Prolific Authors (≥3 Publications)

Author	Number of Publications	Notable Contribution Areas
Denny, P.	7	Computing Education, LLMs for Resource Generation & Feedback
Leinonen, J.	6	Computing Education, LLMs for Code Explanation & Assessment
Bozkurt, A.	4	Generative AI in Education, Systematic Reviews, Ethics
Kasneci, E.	3	Opportunities & Challenges of LLMs in Education
Kuhn, J.	3	AI in Physics Education, Teacher Task Development
Lee, H.	3	ChatGPT in Medical Education
Lo, E.C.M.	3	AI in Dental Education
Luxton-Reilly,	3	Computing Education, AI Code Generation
Prather, J.	3	Computing Education, AI Code Generation, Pedagogy
Tham, Y.-C.	3	AI in Medical Education & Ophthalmology
Ting, D.S.W.	3	Generative AI in Ophthalmology and Medicine
Wong, T.Y.	3	AI in Medical Education & Ophthalmology
Zhai, X.	3	AI in Science Education, Automatic Scoring

The two main authors, Denny and Leinonen, are devoting their main attention to computing education, which is a reflection of the roots of the field. Their research includes automated feedback, exercise generation, and code explanation (Denny et al., 2023; Leinonen et al., 2023; Sarsa et al., 2022). The existence of Bozkurt and Kasneci as authors illustrates the paramount necessity of over-arching frameworks and ethical standards in research. Moreover, the list comprises a good number of medical professionals (Lee, Tham, Ting, Wong), which highlights the strong engagement of applied fields. Collaboration is often very large-scale and within separate clusters (for instance, the computing education cluster around Denny,

3.5 Geographical and Institutional Distribution

The examination of the institutional connections of the corresponding authors reveals the global centers of activity for this research. The United States is the most significant, followed by the United Kingdom, Australia, and China.

Table 5
Top Countries/Regions by Corresponding Author Affiliation

Country/Region	Number of Publications	Country/Region	Number of Publications
United States	58	Canada	8
United Kingdom	18	Singapore	6
Australia	16	South Korea	6
China	15	United Arab Emirates	5
Germany	10	Jordan	4

The US is leading globally, which is noticeable in the fields of AI development and higher education research. UK and Australia collectively produced a significant amount of output that points to the educational technology and medical education research being the main areas of activity in their university sectors. The strong performance of China suggests that it is a major player in educational research and significantly involved in AI applications through education. The participation

of United Arab Emirates and Jordan indicates that the topic is of global interest, often accompanied by an exploration of regional attitudes and adoption practices (Sallam et al., 2023). The University of Auckland (New Zealand) is the most noteworthy institution at the top of the list, owing to the research of Denny and his coworkers. Among the others, the institutions - University of Pennsylvania (US), Harvard Medical School (US), National University of Singapore (Singapore), and University of Cambridge (UK) are the most frequent to appear.

3.6 Thematic Analysis: Key Research Clusters

By means of an investigation into titles, abstracts, and author keywords, the literature can be classified into a number of interconnected thematic clusters. The clusters stand for the main domains of research in the area.

3.6.1 Cluster 1: Benchmarking and Performance Evaluation

This was the preliminary and the simplest line of inquiry. Numerous studies tested the abilities of different LLMs (mainly ChatGPT-3.5, GPT-4, and Google Bard) through professional and academic screening. The findings were continually astonished and indicated that these models were able to score at or above the passing mark for several high-stakes tests.

- **Medical Licensing Exams:** Numerous research works (Gilson et al., 2023; Kung et al., 2023) evaluated the performance on the United States Medical Licensing Examination (USMLE), and also the specialty board exams in neurosurgery (Ali et al., 2023), ophthalmology (Antaki et al., 2023), radiology (Bhayana et al., 2023), orthopaedics (Lum, 2023), and many others.

- **Other Professional Exams:** Chau et al. (2024) carried out similar research in the field of dentistry, while Pursnani et al. (2023) and Dahl et al. (2024) conducted studies in engineering and law, respectively.

- **University-Level Assessments:** Performance was also evaluated through university-level examinations in programming (Savelka et al., 2023), physics (Polverini & Gregorcic, 2024), and chemical engineering (Tsai et al., 2023).

The joint findings indicate that LLM models hold a formidable store of knowledge and reasoning power, thereby challenging conventional notions of valid assessment in the AI era.

3.6.2 Cluster 2: Academic Integrity, Assessment, and Pedagogy

One of the significant events that LLMs' ability to produce very human-like text by the whole quality scale, made it to the forefront of discussions about academic integrity and the issue of cheating and was regarded as one of the most critical arguments against AI, (Perkins, 2023). This grouping contains conversations and inquiries into:

- **Cheating and Plagiarism:** The students' possible options for using LLMs in an unethical manner to finish their assignments and the problems related to the detection of such activities are examined (Susnjak & McIntosh, 2024).

- **Assessment Redesign:** It is suggested that instead of take-home essays and conventional examinations, in-person assessments, oral examinations, project-based learning, and new formats that incorporate AI use would be the right choice, for example "prompt problems" (Denny et al., 2024) and AI-augmented tasks.

- **Pedagogical Transformation:** There has been a significant change in the education process that is needed, pointing out that the new skills such as critical thinking, AI literacy, prompt engineering and source evaluation deserve priority over rote memory (Cain, 2024; Knoth et al., 2024). The term "human-AI collaboration" got to be the central idea, thus classifying the LLM as a means to upgrade, not substitute, human intelligence (Jeon & Lee, 2023).

3.6.3 Cluster 3: Applications in Teaching and Learning

Even though there are risks associated with LLMs, there has been some positive research which has investigated the ways in which LLMs can be used in the education sector as well as their impact on education.

- **Personalized Learning and Tutoring:** LLMs can be harnessed as around-the-clock instructors that, depending on each specific learner's requirements, can respond to queries, clarify doubts, and even create personalized exercises (Sajja et al., 2024; Pesovski et al., 2024).

- **Automated Feedback and Evaluation:** Using LLMs to give developmental criticism on student compositions (Escalante et al., 2023; Guo & Wang, 2024) and to fully automate the evaluation of open-ended answers (Lee, G.-G. et al., 2024; Latif & Zhai, 2024) are some other applications of such technology.

- **Content and Resource Generation:** Helping educators in building lesson plans, syllabi, quiz questions, and illustrative examples (Sarsa et al., 2022; Küchemann et al., 2023).

- **Language Education:** One of the most important sub-clusters is based on the application of ChatGPT for language learning, and this includes writing help, vocabulary enlargement, and dialogue practice (Kostka & Toncelli, 2023; Law, 2024; Zhang & Huang, 2024).

3.6.4 Cluster 4: Medical and Healthcare Education

According to the journal's information, this is a very large and unique group. It is not limited to just comparing performance but also goes on to see areas of use and problems with medical education:

- **Patient Education and Communication:** El LLM se está utilizando para el proceso of translating medical jargon into common language in discharge summaries (Zaretsky et al., 2024) or patient information leaflets (Cheong et al., 2024; Moons & Van Bulck, 2024) as well.

- **Clinical Reasoning and Diagnostic Support:** Structure_LLLMs_ investigated for the treatment of clinical decision-making and differential diagnosis (Tripathi et al., (2014)).

- **Medical Simulation:** The study covered by Holderried et al. (2024) used chatbots in the training of their history-taking simulation; the procedure is integral for students to be able to learn how to handle those of this special patient population.

- **Research and Documentation:** Assistance could be sought for systematic review. Research grants may additionally be made available. Domains of interest are medical writing, research report writing, and administrative work (Gupta et al., 2023).

3.6.5 Cluster 5: Ethics, Limitations, and Future Directions

A recurring theme in the literature is unequivocal emphasis on the limitations and risks of LLMs.

- **Hallucinations and Inaccuracy:** LLMs have been constantly pointed out to possess the propensity to create information that is very believable but at the same time not true or made-up, which is considered a huge risk in areas such as medicine (Miao et al., 2024).

- **Bias and Fairness:** Research has indicated that large language models (LLMs) can not only maintain but also magnify social biases that are associated with race, gender, and culture (Zack et al., 2024).

- **Privacy and Security:** Fear regarding the entrance of particularly sensitive student or patient data into the hands of corporate SLLM platforms.

- **Access and Equity:** In the event that access to cutting-edge AI tools is considered a major factor in determining educational success, the "digital divide" might grow wider (Wang, X. et al., 2023).

- **Transparency and Accountability:** Once it comes to rendering LLMs accountable for their decisions, the problem stems from the "black box" that these models represent.

3.7 Keyword Analysis

An analysis of author keywords confirms the thematic clusters determined above. Table 6 presents the most frequently hapenning keywords (excluding the search terms “large language models” and “education”). The massive prevalence of "ChatGPT" (152 times) highlights its importance as the driver and main research subject in this era. The technology's conceptual framing is also signaled by the powerful presence of "Generative AI". "Medical Education" and "Higher Education" are the specific fields where this research is most active. The rise of "Prompt Engineering" as a common keyword implies that the ability to communicate effectively with LLMs is now considered a new and necessary skill.

Table 6
Most Frequent Author Keywords

Keyword	Frequency	Keyword	Frequency
ChatGPT	152	Natural Language Processing	12
Artificial Intelligence	52	Academic Integrity	10
Generative AI / Generative Artificial	35	Prompt Engineering	9
Medical Education	22	Assessment	8
Higher Education	14	Healthcare	7
AI in Education	13	Ethics	6

4. Discussion: Synthesis of Opportunities and Challenges

The scientometric analysis discloses a field characterized by a dialectic between immense potential and significant peril. The literature oscillates between these two poles, synthesized as follows.

4.1 Predominant Opportunities

1. **Democratization of Tutoring and Support:** LLMs may give personalized, instantaneous academic support to graduate students, potentially leveling the playing field by offering a resource that is not restricted by institutional funding or office hours (Sajja et al., 2024).

2. **Enhancement of Educator Efficiency:** With the help of LLMs, not only generating quiz questions, giving initial feedback on writing, and creating teaching materials are performed as routine tasks but also educators could concentrate on activities such as mentorship, discussion facilitation, and complex problem-solving which require higher-order thinking (Sarsa et al., 2022).

3. **Innovation in Assessment and Pedagogy:** The risk posed by traditional assessment methods is, at the same time, a chance to create more genuine, inventive, and pertinent evaluation forms that would include such abilities as critical thinking, teamwork, and AI-assisted problem-solving among others, as their parameters for measurement (Denny et al., 2024; Eager & Brunton, 2023).

4. **Bridging Theory and Practice in Professional Training:** In areas such as health care and legal practice, LLMs can provide more realistic simulations, propose multi-faceted case studies, and interpret expert knowledge into workable plans, thus improving clinical and practical reasoning skills (Holderried et al., 2024; Tripathi et al., 2024).

4.2 Predominant Challenges

1. **Erosion of Academic Integrity:** The ability to easily use LLMs to generate decent work directly impacts the authenticity of take-home assignments; and it could be a source for creating a culture of outsourcing learning (Perkins, 2023; Susnjak & McIntosh, 2024).

2. **Reliability and "Hallucinations":** The very nature of LLMs to produce incorrect but certain statements has turned them into unreliable sources of information in the first place. This gives rise to the necessity for strong AI literacy including capabilities in verification and critical evaluation (Miao et al., 2024).

3. **Embedded Biases:** The LLMs' training data mirrors the internet's biases, thus making them able to continue stereotypes and create content that is unfair or discriminatory, which is extremely harmful in the domains of education and healthcare, for instance (Zack et al., 2024).

4. **Deskilling and Over-Reliance:** Yet, the risk of falling into the death grip of LLMs has been articulated in terms of basic skills: writing, coding, critical analysis, and the ability to think from first principles (Crawford et al., 2024).

5. **Privacy and Intellectual Property:** The utility of student data for model improvement and the copyright status of AI-generated materials constitute legally ambiguous matters.

5. Conclusion and Future Directions

This scientometric analysis of 200 most-cited publications has drawn a precise picture of the original academic context of LLMs in education. In no time, the discipline underwent transformation from non-existence to a large literature, which was only possible through the disruptive power of ChatGPT. The studies have taken place worldwide, been based on different disciplines, and have involved very close collaborations, among which one of the dominant areas was computing and another one was medical education. The first period, which is indicated in this paper, has been mostly diagnostic and reactive: verifying the strengths and weaknesses of the technology, giving warnings about its threats, and coming up with ideas for

its uses. The shifting of the literature towards a more positive and integrative phase, which is reflected in the growing number of research works on teaching tools, curriculum reformation, and ethical codes, is happening now.

Based on the trends identified, several key directions for future research are suggested:

1. **Longitudinal and Ecological Studies:** It is essential to conduct long-term research that not only focuses on single experiments but also looks into the upholding effect of the incorporation of LLM on the aforementioned areas of student learning, engagement, and skill development in authentic classroom scenarios.

2. **Teacher Professional Development:** Research ought to prioritize the establishment of successful models for the training of teachers in the appropriate and moral usage of LLMs, improving their prompt engineering and teaching strategies suitable for an AI-assisted classroom.

3. **AI Literacy Curricula:** One of the foremost research-practice areas will be the creation, execution, and assessment of AI literacy curricula that have been standardized along all levels of education for students.

4. **Human-AI Collaborative Learning Models:** In the future, it will be necessary to create and evaluate particular models of human-AI cooperation that maximize the advantages of both, which means that AI will be utilized as a means to develop human intelligence instead of taking over.

5. **Robust Technical and Policy Solutions:** Conducting research is obligatory for the introduction of more dependable techniques (if necessary) for spotting AI-generated content, alleviating model biases, and guaranteeing data privacy, in addition to the establishment of legislative and institutional policies for the ethical application of AI in education.

The process of incorporating large language models in the educational sector has only started. Through this survey, the worldwide research community expresses its first and most intensive reactions in various ways. The road ahead will demand unceasing alertness, the application of novel ideas, and an unwavering allegiance to the fundamental educational principles of reasoned discourse, uprightness, and universal access to information.

References

Abd-Alrazaq, A., AlSaad, R., Alhuwail, D., Ahmed, A., Healy, P. M., Latifi, S., Aziz, S., Damseh, R., Alrazak, S. A., & Sheikh, J. (2023). Large language models in medical education: Opportunities, challenges, and future directions. *JMIR Medical Education, 9*, e48291. https://doi.org/10.2196/48291

Ali, R., Tang, O. Y., Connolly, I. D., Fridley, J. S., Shin, J. H., Zadnik Sullivan, P. L., Cielo, D., Oyelese, A. A., Doberstein, C. E., Telfeian, A. E., Gokaslan, Z. L., & Asaad, W. F. (2023). Performance of ChatGPT, GPT-4, and Google Bard on a neurosurgery oral boards preparation question bank. *Neurosurgery, 93*(5), 1090–1098. https://doi.org/10.1227/neu.0000000000002551

Antaki, F., Touma, S., Milad, D., El-Khoury, J., & Duval, R. (2023). Evaluating the performance of ChatGPT in ophthalmology: An analysis of its successes and shortcomings. *Ophthalmology Science, 3*(4), 100324. https://doi.org/10.1016/j.xops.2023.100324

Bhayana, R., Krishna, S., & Bleakney, R. R. (2023). Performance of ChatGPT on a radiology board-style examination: Insights into current strengths and limitations. *Radiology, 307*(5), e230582. https://doi.org/10.1148/radiol.230582

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

Bozkurt, A. (2024). GenAI et al.: Cocreation, authorship, ownership, academic ethics and integrity in a time of generative AI. *Open Praxis, 16*(1), 1–10. https://doi.org/10.55982/openpraxis.16.1.654

Cain, W. (2024). Prompting change: Exploring prompt engineering in large language model AI and its potential to transform education. *TechTrends, 68*(1), 47–57. https://doi.org/10.1007/s11528-023-00896-0

Chau, R. C. W., Thu, K. M., Yu, O. Y., Hsung, R. T.-C., Lo, E. C. M., & Lam, W. Y. H. (2024). Performance of generative artificial intelligence in dental licensing examinations. *International Dental Journal, 74*(3), 616–621. https://doi.org/10.1016/j.identj.2023.12.007

Cheong, R. C. T., Unadkat, S., Mcneillis, V., Williamson, A., Joseph, J., Randhawa, P., Andrews, P., & Paleri, V. (2024). Artificial intelligence chatbots as sources of patient education material for obstructive sleep apnoea: ChatGPT versus Google Bard. European Archives of Oto-Rhino-Laryngology, 281(2), 985–993. https://doi.org/10.1007/s00405-023-08319-9

Crawford, J., Cowling, M., & Allen, K.-A. (2023). Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching and Learning Practice, 20*(3). https://doi.org/10.53761/1.20.3.02

Crawford, J., Allen, K.-A., Pani, B., & Cowling, M. (2024). When artificial intelligence substitutes humans in higher education: The cost of loneliness, student success, and retention. *Studies in Higher Education, 49*(5), 883–897. https://doi.org/10.1080/03075079.2024.2326956

Dahl, M., Magesh, V., Suzgun, M., & Ho, D. E. (2024). Large legal fictions: Profiling legal hallucinations in large language models. *Journal of Legal Analysis, 16*(1), 64–93. https://doi.org/10.1093/jla/laae003

Denny, P., Kumar, V., & Giacaman, N. (2023). Conversing with Copilot: Exploring prompt engineering for solving CS1 problems using natural language. In SIGCSE 2023 - Proceedings of the 54th ACM Technical Symposium on Computer Science Education (Vol. 1, pp. 1136–1142). https://doi.org/10.1145/3545945.3569823

Denny, P., Leinonen, J., Prather, J., Luxton-Reilly, A., Amarouche, T., Becker, B. A., & Reeves, B. N. (2024). Prompt problems: A new programming exercise for the generative AI era. In SIGCSE 2024 - Proceedings of the 55th ACM Technical Symposium on Computer Science Education (Vol. 1, pp. 296–302). https://doi.org/10.1145/3626252.3630909

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eager, B., & Brunton, R. (2023). Prompting higher education towards AI-augmented teaching and learning practice. *Journal of University Teaching and Learning Practice, 20*(5). https://doi.org/10.53761/1.20.5.02

Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics, 105*(3), 1809–1831. https://doi.org/10.1007/s11192-015-1645-z

Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education, 20*(1), 57. https://doi.org/10.1186/s41239-023-00425-2

Eysenbach, G. (2023). The role of ChatGPT, generative language models, and artificial intelligence in medical education: A conversation with ChatGPT and a call for papers. *JMIR Medical Education, 9*, e46885. https://doi.org/10.2196/46885

Gilson, A., Safranek, C. W., Huang, T., Socrates, V., Chi, L., Taylor, R. A., & Chartash, D. (2023). How does ChatGPT perform on the United States Medical Licensing Examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education, 9*, e45312. https://doi.org/10.2196/45312

Guo, K., & Wang, D. (2024). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies, 29*(7), 8435–8463. https://doi.org/10.1007/s10639-023-12146-0

Gupta, R., Herzog, I., Weisberger, J., Chao, J., Chaiyasate, K., & Lee, E. S. (2023). Utilization of ChatGPT for plastic surgery research: Friend or foe? *Journal of Plastic, Reconstructive and Aesthetic Surgery, 80*, 145–147. https://doi.org/10.1016/j.bjps.2023.03.004

Holderried, F., Stegemann-Philipps, C., Herschbach, L., Moldt, J.-A., Nevins, A., Griewatz, J., Holderried, M., Herrmann-Werner, A., Festl-Wietek, T., & Mahling, M. (2024). A generative pretrained transformer (GPT)-powered chatbot as a simulated patient to practice history taking: Prospective, mixed methods study. *JMIR Medical Education, 10*(1), e53961. https://doi.org/10.2196/53961

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, 28*(12), 15873–15892. https://doi.org/10.1007/s10639-023-11834-1

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence, 6*, 100225. https://doi.org/10.1016/j.caeai.2024.100225

Kostka, I., & Toncelli, R. (2023). Exploring applications of ChatGPT to English language teaching: Opportunities, challenges, and recommendations. TESL-EJ, 27(3). https://doi.org/10.55593/ej.27107int

Küchemann, S., Steinert, S., Revenga, N., Schweinberger, M., Dinc, Y., Avila, K. E., & Kuhn, J. (2023). Can ChatGPT support prospective teachers in physics task development? *Physical Review Physics Education Research, 19*(2), 020128. https://doi.org/10.1103/PhysRevPhysEducRes.19.020128

Kung, T. H., Cheatham, M., Medenilla, A., Sillos, C., De Leon, L., Elepaño, C., Madriaga, M., Aggabao, R., Diaz-Candido, G., Maningo, J., & Tseng, V. (2023). Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digital Health, 2*(2), e0000198. https://doi.org/10.1371/journal.pdig.0000198

Latif, E., & Zhai, X. (2024). Fine-tuning ChatGPT for automatic scoring. *Computers and Education: Artificial Intelligence, 6*, 100210. https://doi.org/10.1016/j.caeai.2024.100210

Law, L. (2024). Application of generative artificial intelligence (GenAI) in language teaching and learning: A scoping literature review. *Computers and Education Open, 6*, 100174. https://doi.org/10.1016/j.caeo.2024.100174

Lee, G.-G., Latif, E., Wu, X., Liu, N., & Zhai, X. (2024). Applying large language models and chain-of-thought for automatic scoring. *Computers and Education: Artificial Intelligence, 6*, 100213. https://doi.org/10.1016/j.caeai.2024.100213

Leinonen, J., Denny, P., Macneil, S., Sarsa, S., Bernstein, S., Kim, J., & Tran, A. (2023). Comparing code explanations created by students and large language models. In *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE* (Vol. 1, pp. 124–130). https://doi.org/10.1145/3587102.3588785

Lum, Z. C. (2023). Can artificial intelligence pass the American Board of Orthopaedic Surgery Examination? Orthopaedic residents versus ChatGPT. *Clinical Orthopaedics and Related Research, 481*(8), 1623–1630. https://doi.org/10.1097/CORR.0000000000002704

MacNeil, S., Tran, A., Mogil, D., Bernstein, S., Ross, E., & Huang, Z. (2022). Generating diverse code explanations using the GPT-3 large language model. In ICER 2022 - Proceedings of the 2022 ACM Conference on International Computing Education Research (Vol. 2, pp. 37–39). https://doi.org/10.1145/3501709.3544280

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., & Moore, J. H. (2023). ChatGPT and large language models in academia: Opportunities and challenges. *BioData Mining, 16*(1), 20. https://doi.org/10.1186/s13040-023-00339-9

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence, 6*, 100199. https://doi.org/10.1016/j.caeai.2023.100199

Miao, J., Thongprayoon, C., Suppadungsuk, S., Garcia Valencia, O. A., & Cheungpasitporn, W. (2024). Integrating retrieval-augmented generation with large language models in nephrology: Advancing practical applications. *Medicina (Lithuania), 60*(3), 445. https://doi.org/10.3390/medicina60030445

Moons, P., & Van Bulck, L. (2024). Using ChatGPT and Google Bard to improve the readability of written patient information: A proof of concept. *European Journal of Cardiovascular Nursing, 23*(2), 122–126. https://doi.org/10.1093/eurjcn/zvad087

Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice, 20*(2), 07. https://doi.org/10.53761/1.20.02.07

Pesovski, I., Santos, R., Henriques, R., & Trajkovik, V. (2024). Generative AI for customizable learning experiences. *Sustainability (Switzerland), 16*(7), 3034. https://doi.org/10.3390/su16073034

Polverini, G., & Gregorcic, B. (2024). How understanding large language models can inform the use of ChatGPT in physics education. *European Journal of Physics, 45*(2), 025701. https://doi.org/10.1088/1361-6404/ad1420

Pursnani, V., Sermet, Y., Kurt, M., & Demir, I. (2023). Performance of ChatGPT on the US fundamentals of engineering exam: Comprehensive assessment of proficiency and potential implications for professional environmental engineering practice. *Computers and Education: Artificial Intelligence, 5*, 100183. https://doi.org/10.1016/j.caeai.2023.100183

Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. *Information (Switzerland), 15*(10), 596. https://doi.org/10.3390/info15100596

Sallam, M. (2023). ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare (Switzerland), 11*(6), 887. https://doi.org/10.3390/healthcare11060887

Sallam, M., Salim, N. A., Barakat, M., Al-Mahzoum, K., Al-Tammemi, A. B., Malaeb, D., Hallit, R., & Hallit, S. (2023). Assessing health students' attitudes and usage of ChatGPT in Jordan: Validation study. *JMIR Medical Education, 9*(1), e48254. https://doi.org/10.2196/48254

Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022). Automatic generation of programming exercises and code explanations using large language models. In ICER 2022 - Proceedings of the 2022 ACM Conference on International Computing Education Research (Vol. 1, pp. 27–43). https://doi.org/10.1145/3501385.3543957

Savelka, J., Agarwal, A., An, M., Bogart, C., & Sakr, M. (2023). Thrilled by your progress! Large language models (GPT-4) no longer struggle to pass assessments in higher education programming courses. In ICER 2023 - Proceedings of the 2023 ACM Conference on International Computing Education Research (Vol. 1, pp. 78–92). https://doi.org/10.1145/3568813.3600142

Susnjak, T., & McIntosh, T. R. (2024). ChatGPT: The end of online exam integrity? *Education Sciences, 14*(6), 656. https://doi.org/10.3390/educsci14060656

Tripathi, S., Sukumaran, R., & Cook, T. S. (2024). Efficient healthcare with large language models: Optimizing clinical workflow and enhancing patient care. *Journal of the American Medical Informatics Association, 31*(6), 1436–1440. https://doi.org/10.1093/jamia/ocad258

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. https://doi.org/10.1016/j.ece.2023.05.001

Wang, X., Sanders, H. M., Liu, Y., Seang, K., Tran, B. X., Atanasov, A. G., Qiu, Y., Tang, S., Car, J., Wang, Y. X., Wong, T. Y., Tham, Y.-C., & Chung, K. C. (2023). ChatGPT: Promise and challenges for deployment in low- and middle-income countries. The Lancet Regional Health - Western Pacific, 41, 100905. https://doi.org/10.1016/j.lanwpc.2023.100905

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology, 55*(1), 90–112. https://doi.org/10.1111/bjet.13370

Zack, T., Lehman, E., Suzgun, M., Rodriguez, J. A., Celi, L. A., Gichoya, J., Jurafsky, D., Szolovits, P., Bates, D. W., Abdulnour, R.-E. E., Butte, A. J., & Alsentzer, E. (2024). Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: A model evaluation study. *The Lancet Digital Health, 6*(1), e12–e22. https://doi.org/10.1016/S2589-7500(23)00225-X

Zaretsky, J., Kim, J. M., Baskharoun, S., Zhao, Y., Austrian, J., Aphinyanaphongs, Y., Gupta, R., Blecker, S. B., & Feldman, J. (2024). Generative artificial intelligence to transform inpatient discharge summaries to patient-friendly language and format. *JAMA Network Open, 7*(3), e240357. https://doi.org/10.1001/jamanetworkopen.2024.0357

Zhang, Z., & Huang, X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition. *Heliyon, 10*(3), e25370. https://doi.org/10.1016/j.heliyon.2024.e25370

© 2025 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).