After One Year, Where are Large Language Models Headed: A Thematic Analysis using Bibliometric Methodology

Ethan Bernstein, Anya Ramsamooj, Kelsey Leann Millar, Zachary C Lum

Submitted to: JMIR AI
on: November 10, 2024

Table of Contents

After One Year, Where are Large Language Models Headed: A Thematic Analysis using Bibliometric Methodology

Ethan Bernstein[1] BS; Anya Ramsamooj[1] BS; Kelsey Leann Millar[2] MD; Zachary C Lum[2] DO

[1]California Northstate University College of Medicine Elk Grove US
[2]University of California Davis Medical Center Department of Orthopaedic Surgery Sacramento US

Corresponding Author:
Ethan Bernstein BS
California Northstate University
College of Medicine
9700 West Taron Dr
Elk Grove
US

Abstract

Background: Since the release of ChatGPT and other large language models (LLMs), there has been a significant increase in academic publications exploring their capabilities and implications across various fields, such as Medicine, Education, and Technology.

Objective: This study aims to identify the most influential academic works on LLMs published in the past year, categorize their research types and thematic focuses, within different professional fields. The study also evaluates the ability of AI tools, such as ChatGPT, to accurately classify academic research.

Methods: We conducted a bibliometric analysis using Clarivate's Web of Science (WOS) to extract the top 100 most cited articles on LLMs. Articles were manually categorized by field, journal, author, and research type. ChatGPT-4 was used to generate categorizations for the same articles, and its performance was compared to human classifications. Statistical analyses were performed to determine the prevalence of research fields and to evaluate the accuracy of AI-generated classifications.

Results: Medicine emerged as the predominant field among the top-cited articles, accounting for 43%, followed by Education (26%) and Technology (15%). Medical literature primarily focused on clinical applications of LLMs, limitations of AI in healthcare, and the role of AI in medical education. In Education, research was centered around ethical concerns and potential applications of AI for teaching and learning. ChatGPT demonstrated high concordance with human reviewers, achieving an agreement rating of 86% for research types and 92% for fields of study.

Conclusions: While LLMs like ChatGPT exhibit considerable potential in aiding research categorization, human oversight remains essential to address issues such as hallucinations, outdated information, and biases in AI-generated outputs. This study highlights the transformative potential of LLMs across multiple sectors and emphasizes the importance of continuous ethical evaluation and iterative improvement of AI systems to maximize their benefits while minimizing risks.

(JMIR Preprints 10/11/2024:68603)
DOI: https://doi.org/10.2196/preprints.68603

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

After One Year, Where are Large Language Models Headed: A Thematic Analysis using Bibliometric Methodology

Authors: Ethan M. Bernstein, BS[1], Anya Ramsamooj, BS[1], Kelsey L. Millar, MD[2], Zachary C. Lum, DO[2]

[1]California Northstate University College of Medicine, Elk Grove CA, [2]Department of Orthopaedic Surgery, University of California Davis Medical Center, Sacramento, CA

Corresponding Author: Ethan Bernstein, BS

Email: emdb711@gmail.com

Phone: (847) 917 - 5890

Institution Address: 9700 W Taron Dr, Elk Grove, CA 95757

Phone: (916) 378 - 3491

Fax: (916) 686 – 7310

Word Count: 4006

Abstract

Background: Since the release of ChatGPT and other large language models (LLMs), there has been a significant increase in academic publications exploring their capabilities and implications across various fields, such as Medicine, Education, and Technology.

Objective: This study aims to identify the most influential academic works on LLMs published in the past year, categorize their research types and thematic focuses, within different professional fields. The study also evaluates the ability of AI tools, such as ChatGPT, to accurately classify academic research.

Methods: We conducted a bibliometric analysis using Clarivate's Web of Science (WOS) to extract the top 100 most cited articles on LLMs. Articles were manually categorized by field, journal, author, and research type. ChatGPT-4 was used to generate categorizations for the same articles, and its performance was compared to human classifications. Statistical analyses were performed to determine the prevalence of research fields and to evaluate the accuracy of AI-generated classifications.

Results: Medicine emerged as the predominant field among the top-cited articles, accounting for 43%, followed by Education (26%) and Technology (15%). Medical literature primarily focused on clinical applications of LLMs, limitations of AI in healthcare, and the role of AI in medical education. In Education, research was centered around ethical concerns and potential applications of AI for teaching and learning. ChatGPT demonstrated high concordance with human reviewers, achieving an agreement rating of 86% for research types and 92% for fields of study.

Conclusion: While LLMs like ChatGPT exhibit considerable potential in aiding research categorization, human oversight remains essential to address issues such as hallucinations, outdated information, and biases in AI-generated outputs. This study highlights the transformative potential of LLMs across multiple sectors and emphasizes the importance of continuous ethical evaluation and iterative improvement of AI systems to maximize their benefits while minimizing risks.

Keywords: Large Language Models; ChatGPT; Web of Science; Medicine; Education; Technology ;Research Categorization; Artificial Intelligence.

Level of Evidence: III

Introduction

Within academic literature, artificial intelligence (AI) is broadly defined as a mechanical emulation of the human thinking processes to facilitate the analysis, simulation exploitation, and exploration of

human thinking processes.[1] Chat Generative Pre-trained Transformer (ChatGPT) has been trained on a massive amount of internet data from 2021 and is being updated from here on to include current data and information retrieval from the internet, which should reflect current, up to date information. It utilizes deep neural networks, machine learning, and the training data to interact with prompts and generate relevant human-like text responses,[2] qualifying it as a "Large Language Model" (LLM).[3] ChatGPT may have significant potential to improve the efficiency of human innovation in a multitude of fields that require quick access to information, such as Medicine and Education, by providing instant feedback and helping to expedite clerical work, such as writing notes or research processes.

It has been over a year since the public release of ChatGPT by the company OpenAI (San Francisco, CA, USA), followed by Gemini (formerly BARD) by Google (Mountainview, CA, USA) and subsequent widespread use of AI, from usage by students for writing support to researchers for literature review and manuscript preparation assistance. From its release through August 2023, there have been over 1000 PubMed citations,[4] indicating its rapid rise in utilization and interest in its capabilities. With any new technology, many use cases are developed and tested, until a narrowing of the field emerges. This determines which aspects of the technology resonate within the community and which aspects need further refinement. We sought to determine which types of studies were performed, which fields of research have the most studies, and journals are the most highly cited, as this may highlight fields that are the most intense focus of researchers regarding the use of AI. We also sought to determine how well ChatGPT performs thematic categorization of research type and field of research when analyzing academic publications.

Methods

We used Clarivate Web of Science (WOS) and searched for all research articles with the terms "chatgpt" and "bard" and "large language model" independently in March 2024. Articles with the 100 highest citation counts were exported into a spreadsheet generated from the search results

through WOS, which were then categorized by journal, author, type of research, and field of research by one independent author (E.B). We also used ChatGPT-4 (GPT-4 model) to thematically categorize them and compare manual grouping to AI. The literature review was conducted from June to July 2024.

Author's Categorization

Python (version 3.11.7) was used to count the number of times an author was listed in the top 100 cited publications in the Excel spreadsheet generated by WOS and rank the authors according to frequency. Python was also used to count the number of times a journal was listed in the top 100 cited publications and rank in order of frequency. This code is depicted in Appendix S2a-b. Based on review of each study's abstract, the papers were categorized by the general field of research that the study was conducted in, and the type of research that was conducted - counted manually in an Excel spreadsheet. Some papers could be categorized under the purview of multiple fields; for example, papers discussing medical education could have been categorized as either Medicine or Education. In such cases, the primary field of focus of the publishing journal was referred to for reconciliation. These tasks were carried out by a single author (E.B). The type of research was determined by reading the methods of each study and recorded, and Python was used to count and sort by frequency (Appendix S2c). During the literature review phase, the field of study and research type was either confirmed or changed for each paper. The code used is depicted in Appendix S2.

Literature Review

Following the categorization stage, we determined the 3 most cited areas of research and a full literature review was performed of the included works. Top 3 areas were determined by which field had the most papers listed in the top 100 based on the authors determination of field of study. In total, there were 84 articles included in the literature review. Each paper was analyzed to determine the primary topics of interest regarding AI and LLMs. A list of topics was generated as each paper

was read, and the number of articles that covered each topic were counted in a spreadsheet. For example, if a paper in the medicine category covered the Clinical Uses of ChatGPT, it was marked and counted towards the total number of articles that covered that topic. This was done by both authors (E.B, A.R) performing the literature review and any discrepancies were reconciled by discussion between the 2 reviewers. These areas are further discussed in the Results section of this paper. The goal of this study is to explore the fields that are most interested in AI LLMs, and assess current and future implications of AI LLMs in each respective field. Additionally, we aimed to and evaluate ChatGPT's ability to analyze and categorize scientific literature.

Thematic Categorization Performed by ChatGPT (GPT-4)

ChatGPT (GPT-4 model) was asked to count and report authors and journals in order of frequency. The list of authors and journals was pasted into ChatGPT, and ChatGPT was prompted to sort by frequency. These generated lists were then compared to the corresponding lists generated by the authors. To determine the study type and field of research, the PDF of each paper was imported into ChatGPT, which was asked to determine study type and field of research. These results were then directly compared to the corresponding author generated results and assessed for accuracy. Examples of these prompts are listed in Appendix S3.

Results

Publication Sources

Overall, there were 12 authors whose names appeared twice in the top 100 citations, but none appeared more than twice. Across the top 100 cited papers, the Cureus Journal of Medical Science had the most, with seven papers, while Education Sciences and Journal of Internet Research each had three papers. Every other journal had either one or two publications on the top 100 cited papers list (Appendix S4).

Field of Research

Medicine was the most cited area of study with 43 papers, followed by Education and Educational Research with 26, followed by Technology and Information Sciences with 15 papers. Business and Economics had 4, and Tourism had 4. All other areas had either 2 or 1 paper in the top 100. These findings are summarized in Table 1 with further breakdown of individual categorizations in Appendices S5-7.

Table 1. Areas of research as generated by the author. *=Top 3 area of research included in literature review.

Area of Study	**Number of publications (n=100)**
Medicine	43*
Education	26*
Technology and Information Sciences	15*
Business and Economics	4
Tourism	4
Government/Law	2
Public Health	2
Basic Sciences	1
Ethics	1
Geography	1
Pharmacology	1

Multiple medical subfields published papers regarding ChatGPT. General and Internal Medicine was the most mentioned subcategory with 11 publications, followed by Health Care Sciences and Services with 7, Surgery with 5, Oncology with 4, Radiology with 3, and Ophthalmology with 3. All other areas of medicine had only 1 publication mentioned. These findings are summarized in Table 2.

Table 2. Breakdown of specific fields of medicine that papers were published in.

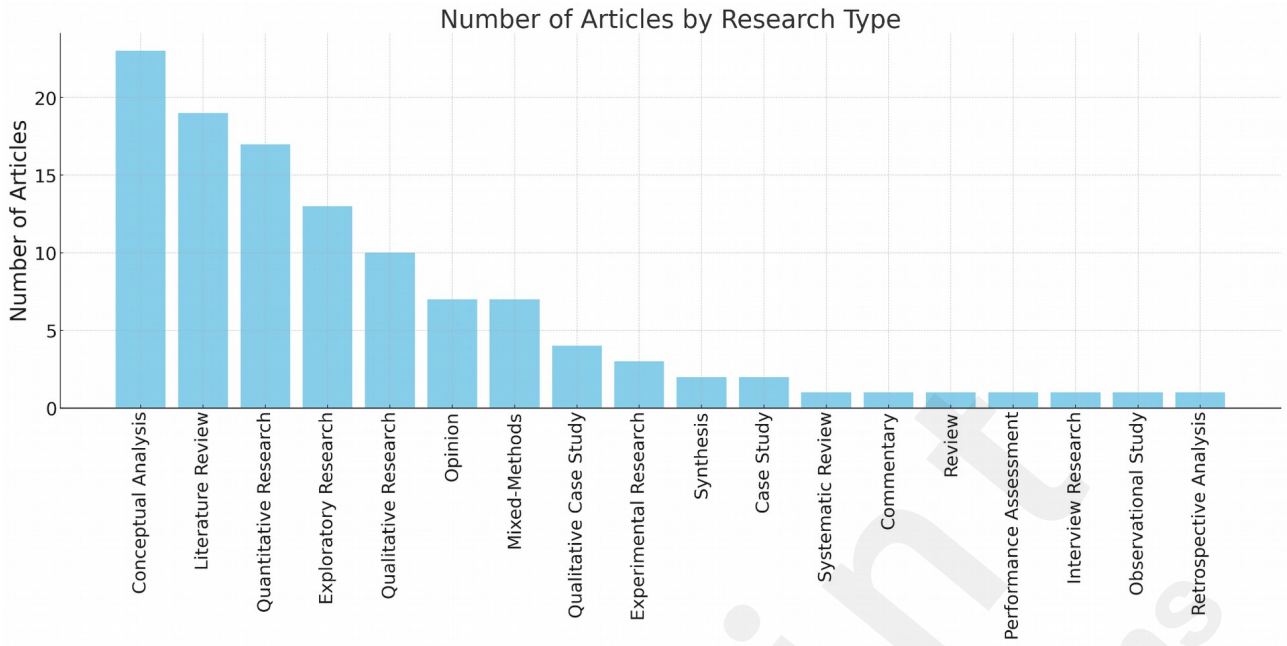
Field of Medicine	**# Of Articles**
General and Internal Medicine	11
Health Care Sciences and Services	8
Surgery	5
Oncology	4

Radiology, Nuclear Medicine & Medical Imaging	3
Ophthalmology	3
Nursing	1
Gastroenterology and Hepatology	1
Endocrinology and Metabolism	1
Otorhinolaryngology	1
Medical Ethics	1
Sport Sciences	1
Dentistry, Oral Surgery and Medicine	1
Biomedical Social Sciences	1
Nursing	1
Total # of Papers in Medicine	43

Research Type

The most common research type was conceptual analysis, which included papers exploring theoretical applications of AI models without conducting any experiments. 23 papers were categorized as conceptual analysis, determined by the author (E.B). Literature review was the next most frequent study type with 18 papers, followed by quantitative research at 17, exploratory research with 12, and qualitative research with 11. The rest of the research types are depicted in Figure 1. Many of the papers summarized theoretical uses, as well as potential future uses of ChatGPT, that are still being explored.

Figure 1. Bar graph depicting the study type frequency of the included publications.



After uploading the PDF copy of each paper, ChatGPT was asked to identify the field of research and study type. In instances of discordant answers for fields of research or research types, the author (E.B.) evaluated ChatGPT's answer to determine plausibility. In evaluating the agreement between research types between the author and ChatGPT, the percentage agreement was found to be 86%. The percentage agreement for the field of research was 92%.

Medicine

Of the 43 papers[4–42] reviewed from the field of medicine, 38 (84.44%) of the papers discussed the limitations of ChatGPT, 30 (66.67%) discussed the uses of ChatGPT in clinical medicine, 21 (46.67%) numerically evaluated the quality and capability of ChatGPT in regards to medical reasoning, and nine (20%) evaluated the uses of ChatGPT in medical research. Medical education was an additional area of focus with 14 (32.56%) papers discussing ChatGPT's ability to answer board certification questions or help with studying from both the student and educator perspective. These findings are summarized in Table 3.

Table 3. The following table provides a comprehensive breakdown of various topics covered by existing research in medicine. This table offers a quantitative overview, highlighting the predominant focus areas and revealing gaps that might need further exploration.

Topic	Quality Assessment in Medical Knowledge	Quality Assessment in Research	Limitations of ChatGPT	Uses in Medicine	Uses in Research	Ethics	Medical Education
Sum	**21**	**9**	**38**	**30**	**19**	**18**	**14**
% (n=43)	**46.67%**	**20.00%**	**84.44%**	**66.67%**	**42.22%**	**37.78%**	**31.11%**

Quality assessment in Medical Knowledge most commonly included studies that evaluated the accuracy of ChatGPT's answer to specific medical questions from patients and board examination questions, and the capability to employ higher-order medical reasoning. Many of these papers cited promising results and indicated that further research and technology improvements are necessary prior to full implementation into the medical field. Many papers discussed the already proven or hypothetical uses of ChatGPT in medicine, such as the enhancement of telemedicine, answering patient questions, and administrative tasks such as charting and other paperwork.[5,6] Potential weaknesses preventing current widespread adoption included lack of nuanced information that an experienced physician would understand and potential inaccuracies due biased or outdated training data.[7,8] Additionally, authors cited the high level of confidence that ChatGPT appears to answer with, which may result in the dissemination of misinformation.[9] Further improvement of the technology would be necessary due to lack of deep understanding and inability to interpret complex medical imaging. Some studies compared physician responses to AI responses, and found that evaluators may prefer the AI generated response.

Other uses in medicine included assistance for non-native English speakers with translation to their native language.[10–12] AI is able to quickly comprehend information, which indicates a potential use for providing information about medical guidelines in acute situations, such as in the ICU.[5] AI may

also be useful in assisting with documentation, decision support, and patient communication.

Quality assessment in research discussed automatic generation of citations, manuscripts, ideas, and hypotheses. The ability of ChatGPT/LLMs to conduct literature searches was also evaluated[33]. Some studies synthesized the already proven uses in research, the most prominent of which included literature review, data synthesis, and assistance with data analysis.[5,43] Review papers covering research utility discussed the ability to generate large amounts of text and can help authors organize their thoughts[16]. ChatGPT has been shown to quickly summarize whole papers well and can save researchers significant amounts of time.[44]

Limitations of AI models was the most common topic covered, with many papers referencing lack of updated information because ChatGPT was trained on data from 2021, which may lead to outdated or incorrect information. The possibility of ChatGPT hallucination, generation of fake citations, and perpetuation of biases contained in the information that it was trained on was also discussed.[10,12] Papers that discussed ethics included privacy concerns with cybersecurity risks with patient confidentiality as ChatGPT stores its conversations in its memory bank with minimal evidence that it can comply with HIPAA regulations.[19] One paper reported concerns with shortcuts in the research or learning process, which could lead to inflated numbers of publications without the same level of expertise.[16]

Papers regarding medical education discussed potential use via interactive simulation, immediate feedback and information, and creation of educational materials for instructors. Students can generate quiz questions, which can alleviate the work burden for educators and students alike. Studies such as Huh et al. 2023 showed that ChatGPT performed comparably to medical students in certain assessments indicating the potential for integration and utilization in medical education.[45] Multiple studies cited promising results of ChatGPT answering board questions.

Education

26 papers[45,45–70] were categorized as Education as their field of study. 21 (84.61%) Education papers discussed the ethical concerns associated with AI implementation in schools; 7 (26.92%) discussed the potential negative implications on students' learning and performance capabilities, and 17 (65.38%) discussed solutions for these concerns. 18 (69.23%) papers discussed the potential implementations for students and 17 (65.38%) discussed potential implementations for educators. These findings and the other findings are summarized in Table 4.

Table 4. The following table provides a comprehensive breakdown of various topics covered by existing research in education.

Topic	Privacy	Ethics	Negative effects on students	Solutions for AI Concerns	Quality assessment	Capabilities of AI	Uses for students	Uses for Educators	Limitations
Sum	3	21	7	17	9	13	18	17	18
% (n=26)	12%	81%	27%	65%	35%	50%	69%	65%	69%

In the field of Education, the main privacy concerns were related to the storage of student data and personal information that may be stored by ChatGPT through regular usage. Papers discussed the importance of data security and compliance with privacy regulations to prevent personal information being disseminated.

Education ethics primarily discussed academic dishonesty including plagiarism and incorrect citations. Evidence advocating for ChatGPT as a reliable author and/or primary source is minimal, and the academic ethical implications of copying and pasting the ChatGPT response was unclear; however, the consensus questioned the ethics of directly quoting ChatGPT without attributing it as the source.[46,71] Additional sources of ethical concerns were regarding the discordant access to ChatGPT, as the optimized version is currently a paid service.[47,48] Additionally, perpetuation of discriminatory and biased ideas in the information that ChatGPT was trained on was a common

concern, which led papers to recommend cautious use of AI and special care to identify and mitigate such biases.[17,49] Educators have expressed concerns over students including overreliance leading to loss of writing skills and hindrance of creativity in addition to academic dishonesty.

Solutions to AI overreliance included implementation of AI literacy early in the education system, similar to how typing and technology training became ubiquitous with the rise of computers.[50] Other workarounds to overreliance on ChatGPT include assignment design that is incompatible with ChatGPT.[51,52] Additionally, educators are being urged to specify the permitted usage of ChatGPT in their course syllabi for transparency.[53] On the side of OpenAI, regular updates to ensure accuracy of information available to ChatGPT are also crucial. Perkins et al. 2023 discussed the need for transparency and guidelines for AI use in academic settings.[53]

Quality assessments evaluated accuracy, relevance, and potential biases of text generated by ChatGPT in multiple educational fields generally, or within specific areas like chemistry, language, administration, and academic research. Researchers looked into the capability to enhance learning, teaching and grading. Fergus et al. 2023 evaluated ChatGPT's ability to generate assessment questions in chemistry and the quality of the answers generated by ChatGPT.[54] Farrokhnia et al. 2024 performed a SWOT analysis of ChatGPT in education and research, reporting the ability to provide real-time feedback and grading with specific responses, which can decrease workload for both students and educators.[55] Other papers explored academic integrity and third party program's ability to detect work written by ChatGPT, the ability of ChatGPT to generate educational materials, such as textbooks, study guides, and practice questions, and the ability for AI to assist students with learning experience and motivation. Papers marked under Uses for Students and Uses for Educators included relevant quality assessments and review papers covering proven uses.

ChatGPT's limitations in education include its potential for generating inaccurate or biased information, its dependency on up-to-date training data, and its inability to fully replace human

educators.[58,60] For example, while ChatGPT can provide factual information, it lacks the ability to engage in nuanced discussions or provide emotional support.[48]

Technology and Information Sciences

15 papers[3,43,72–83] were categorized as reviewed Technology and Information Sciences. Subtopics included computer science, engineering and electric vehicles, and data science. 12 (80%) papers discussed the technical uses of ChatGPT, with two of those being quality assessment papers. Six (40%) discussed ethics and privacy concerns with nine (60%) discussing limitations. These findings are summarized in Table 5.

Table 5. The following table provides a comprehensive breakdown of various topics covered by existing research in technology and information sciences.

Topic	Technical uses of ChatGPT	Quality Assessment	Ethics, Privacy Concerns	Public Perception	Limitations	Future
Sum	12	2	6	1	9	9
% (n=15)	80%	13%	40%	7%	60%	60%

Ethics in the field of Technology and Information Sciences included privacy concerns with the handling and storage of user data. Similar to other sections, authors emphasized the importance of strong cybersecurity measures including encryption and secure storage to minimize the risk with security breaches. Uniquely, some papers discussed the potential of AI to create harmful “deep fake” content, which has the potential to be weaponized.[72,84] Deep fake is defined as high-quality fabricated image and video content that can be misinterpreted by viewers as being real.[85]

Technical uses of ChatGPT in this field included code generation, debugging, and automated routine IT processes. The ability to assist with data science through data cleaning, preprocessing, and preliminary result interpretation was explored.[73,74] Du et al. 2023 discussed the potential of AI

implementation in electric vehicles.[75] Holzinger et al. 2023 covered AI from the biotechnology perspective, which provided an overview of the uses of AI to agricultural engineering, medical biotechnology, and bioinformatics.[86] AI is already being utilized by plant tissue scientists to simulate complex interactions and treatment options for their agricultural experimentation. AI can benefit medicine from a micro-perspective, including genomic analysis, biomedical image analysis, data analytics, and drug discovery and development.[86] The advantage comes from the ability to rapidly analyze large quantities of data through automation, and implementation of predictive models that can analyze image data and recommend the best management and planning course for any given task. Quality assessment papers that analyzed potential uses of ChatGPT looked at AI's writing abilities in order to contribute to the conversation surrounding job security.[87]

Limitations of AI in the field of Technology and Information Sciences include difficulty with performance in complex problem-solving scenarios, outdated training data, and potential inaccuracies.[73] Authors emphasized that AI requires human oversight to ensure reliable outputs.[76] For example, while ChatGPT can assist with coding, it might not always understand the context of complex software projects, leading to errors.[81] Future directions that the academic technology community is hoping ChatGPT/AI models take include improvement of privacy concerns, and continued efforts to address ethical concerns.[72] Additionally, improving reliability to decrease the likelihood of hallucination and increase complex understanding to improve usability and reliability.[75]

Finally, one paper examined public perception of ChatGPT, by evaluating responses on Twitter to determine generally how internet users felt about the dawn of AI.[80]

ChatGPT's Ability to Thematically Categorize Author List, Journal, Research Type, and Field of Research

When asked to report the frequency of authors mentioned, ChatGPT generated a different list than the author generated list. It was also unable to correctly count the frequency of each journal. Pictures

of ChatGPT's output are included in appendix S8.

Discussion

This discussion explores the current trajectory of AI integration, potential breakthroughs, and the implications for the following fields.

Medicine

In medicine, AI is primarily being tested to assist in areas from administrative support to clinical applications. The potential of AI to enhance telemedicine, streamline administrative tasks, and assist in diagnostic processes is well-documented.[6,17,49] However, the accuracy and reliability of AI in medical decision-making are areas that require further research. Future pathways include the integration of AI into clinical workflows to assist with patient triage, diagnostic support, and personalized treatment plans.[6] The field appears to be moving towards leveraging AI to augment, rather than replace, human expertise, with potential breakthroughs anticipated in predictive analytics and personalized medicine.[11,33,42,88] The ongoing challenge will be ensuring AI systems are trained on up-to-date and diverse datasets to minimize bias and inaccuracies. Before widespread integration of AI in medicine occurs, further research is necessary to prove reliability and improvement from early models.[17]

Additionally, prior research suggests that programs such as ChatGPT can assist with medical education. For instance, Huh et al. 2023 demonstrated that ChatGPT, while not outperforming medical students, provided reasonable answers on parasitology exams, indicating its potential for educational integration.[45] The literature also indicates that research processes are being disrupted, with AI showing promise in tasks such as automatic generation of citations, manuscripts, and hypotheses, which can streamline the research process.[46] However, reducing the hallucination frequency and accuracy improvements will need to be made before any significant disruption in clinical practice can occur.

Education

Education has the potential to enter a transformative phase with the incorporation of AI tools like ChatGPT into current curricula. These tools can offer substantial benefits, including personalized learning experiences, automated grading, and enhanced teaching aids. However, the potential for academic dishonesty and the ethical implications of AI use in education remain significant concerns.[44,49,50,57,64,67] As AI technology improves, it is expected to offer even more sophisticated support for both students and educators, such as adaptive learning platforms that cater to individual student needs and real-time feedback systems. The field is also exploring AI's role in reducing teacher workload and providing continuous professional development opportunities.[44,47,55,63] Ensuring equitable access to AI tools and addressing ethical dilemmas will be crucial as AI becomes more integrated into educational systems.

Technology and Information Sciences

AI's impact on technology and information sciences, particularly in software development, data analysis, and cybersecurity, has positive indications for progress. AI tools are increasingly used for code generation, something that was successfully implemented in this study to help with counting author, journal, study type frequencies and generating figures. Debugging, automating routine IT tasks, and enhancing productivity and efficiency are other capabilities.[3,78,81] Future breakthroughs are expected in the development of more reliable and context-aware AI systems capable of handling complex problem-solving scenarios. Ethical concerns, such as protecting privacy and security, will continue to be pivotal, with ongoing efforts to develop robust cybersecurity measures and ethical AI guidelines.

Limitations of ChatGPT in Conducting Research

Thematic categorization of the field of research sometimes required information that was not directly

stated in the article, which would make it very difficult for ChatGPT to correctly determine the correct field of research. For example, some papers covered technology topics but were published in a medical journal.[41] ChatGPT would categorize these papers as technology while the author categorized them as medicine. This discrepancy may be attributed to author criteria rather than technology failure; however, it indicates the importance of providing detailed, specific instructions to LLMs in order to receive the desired output. Due to such limitations, it is important for people using ChatGPT for this purpose to monitor the outputs and verify accuracy.

Thematic Categorization

The results of ChatGPT's thematic categorization suggest that further improvements must be made before ChatGPT can reliably sort and organize data. At the time the study was performed, it appeared as though ChatGPT was overwhelmed by large amounts of text, which raises questions about its capabilities of sorting through information such as lists of names and journals; when asked to sort through the list of authors, which contained over 190 names, the sorted list was incorrect. However, it did perform well in determining the research type and field of research, suggesting that ChatGPT can be trusted with simpler and more straightforward tasks.

Conclusion

Medicine, education, and technology are preparing for a future with potential LLM integration, as demonstrated by the high citation counts in these fields. Current research focuses on testing the capabilities of AI models like ChatGPT and disseminating findings to inform best practices. While LLMs show promise in enhancing workflows and categorizing research, this study emphasizes the necessity of human oversight to mitigate risks such as hallucination, outdated information, and biases, which users may trust without cross-referencing. As AI integration evolves, attention to ethical concerns, equitable access, and continuous updates with accurate and up-to-date information will be essential. The potential for AI to transform these sectors can only be fully realized through

responsible, well-managed implementation that augments human expertise and ensures fair use.

In medicine, AI is expected to integrate further into clinical workflows, assisting with diagnostics, patient communication, and administrative tasks like charting, although concerns remain about accuracy and ethical implications. In education, AI tools may be able to revolutionize personalized learning, automate grading, and support educators, but addressing academic dishonesty and ensuring ethical AI use in learning environments is crucial. In technology, LLMs are advancing software development, data science, and cybersecurity, but future work needs to enhance AI's ability to handle complex problem-solving while ensuring privacy and security.

These fields are tightly linked, with educational pedagogy aiding the development of physicians, and the field of technology developing the software and applications that physicians will use. If AI positively affects education, it may result in a network of physicians who can use advanced technology to increase efficiency and improve patient care. Furthermore, AI is being developed to enhance the research process, allowing researchers to be more productive and driving medical discoveries. LLMs are moving toward transforming the efficiency and quality of these fields through continued improvement and integration. Ultimately, the true potential of AI can be realized through a collaborative approach, where human expertise works in tandem with AI, ensuring that both ethics and efficiency are upheld.

Acknowledgements

ChatGPT (GPT-4) was used for grammar and editing assistance for this paper. Thank you to Cameron Bernstein for help with proofreading and editing this paper.

Conflict of Interest

The authors declare that they have no competing interests related to the research, authorship, or

publication of this manuscript.

References

1. Lu Y. Artificial intelligence: a survey on evolution, models, applications and future trends. *Journal of Management Analytics*. 2019;6(1):1-29. doi:10.1080/23270012.2019.1570365
2. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25(1):24-29. doi:10.1038/s41591-018-0316-z
3. Wu T, He S, Liu J, et al. A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development. *IEEE/CAA J Autom Sinica*. 2023;10(5):1122-1136. doi:10.1109/JAS.2023.123618
4. Temsah O, Khan SA, Chaiah Y, et al. Overview of Early ChatGPT's Presence in Medical Literature: Insights From a Hybrid Literature Review by ChatGPT and Human Experts. *Cureus*. Published online April 8, 2023. doi:10.7759/cureus.37281
5. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. *J Med Syst*. 2023;47(1):33. doi:10.1007/s10916-023-01925-4
6. Liu J, Wang C, Liu S. Utility of ChatGPT in Clinical Practice. *J Med Internet Res*. 2023;25:e48568. doi:10.2196/48568
7. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol*. 2023;29(3):721-732. doi:10.3350/cmh.2023.0089
8. Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. *JNCI Cancer Spectrum*. 2023;7(2):pkad015. doi:10.1093/jncics/pkad015
9. Samaan JS, Yeo YH, Rajeev N, et al. Assessing the Accuracy of Responses by the Language Model ChatGPT to Questions Regarding Bariatric Surgery. *OBES SURG*. 2023;33(6):1790-1796. doi:10.1007/s11695-023-06603-5
10. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare*. 2023;11(6):887. doi:10.3390/healthcare11060887
11. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: Current applications, future possibilities and limitations of ChatGPT. *Diagnostic and Interventional Imaging*. 2023;104(6):269-274. doi:10.1016/j.diii.2023.02.003
12. Dergaa I, Chamari K, Zmijewski P, Ben Saad H. From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *bs*. 2023;40(2):615-622. doi:10.5114/biolsport.2023.125623
13. Sinha RK, Deb Roy A, Kumar N, Mondal H. Applicability of ChatGPT in Assisting to Solve Higher Order Problems in Pathology. *Cureus*. Published online February 20, 2023. doi:10.7759/cureus.35237
14. Alkaissi H, McFarlane SI. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus*. Published online February 19, 2023. doi:10.7759/cureus.35179
15. Das D, Kumar N, Longjam LA, et al. Assessing the Capability of ChatGPT in Answering First- and Second-Order Knowledge Questions on Microbiology as per Competency-Based Medical Education Curriculum. *Cureus*. Published online March 12, 2023. doi:10.7759/cureus.36034
16. Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing? *Crit Care*. 2023;27(1):75. doi:10.1186/s13054-023-04380-2
17. Fatani B. ChatGPT for Future Medical and Dental Research. *Cureus*. Published online April 8, 2023. doi:10.7759/cureus.37285
18. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. *Pak J Med Sci*. 2023;39(2). doi:10.12669/pjms.39.2.7653

19. Ayers JW, Poliak A, Dredze M, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. *JAMA Intern Med*. 2023;183(6):589. doi:10.1001/jamainternmed.2023.1838
20. Athaluri SA, Manthena SV, Kesapragada VSRKM, Yarlagadda V, Dave T, Duddumpudi RTS. Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Cureus*. Published online April 11, 2023. doi:10.7759/cureus.37432
21. Waisberg E, Ong J, Masalkhi M, et al. GPT-4: a new era of artificial intelligence in medicine. *Ir J Med Sci*. 2023;192(6):3197-3200. doi:10.1007/s11845-023-03377-8
22. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE. High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content. *Cureus*. Published online May 19, 2023. doi:10.7759/cureus.39238
23. Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened. *J Med Internet Res*. 2023;25:e46924. doi:10.2196/46924
24. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digit Med*. 2023;6(1):75. doi:10.1038/s41746-023-00819-6
25. Choudhury A, Shamszare H. Investigating the Impact of User Trust on the Adoption and Use of ChatGPT: Survey Analysis. *J Med Internet Res*. 2023;25:e47184. doi:10.2196/47184
26. Liu S, Wright AP, Patterson BL, et al. Using AI-generated suggestions from ChatGPT to optimize clinical decision support. *Journal of the American Medical Informatics Association*. 2023;30(7):1237-1245. doi:10.1093/jamia/ocad072
27. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29(8):1930-1940. doi:10.1038/s41591-023-02448-8
28. Will ChatGPT transform healthcare? *Nat Med*. 2023;29(3):505-506. doi:10.1038/s41591-023-02289-5
29. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the Performance of ChatGPT in Ophthalmology. *Ophthalmology Science*. 2023;3(4):100324. doi:10.1016/j.xops.2023.100324
30. Vaishya R, Misra A, Vaish A. ChatGPT: Is this version good for healthcare and research? *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. 2023;17(4):102744. doi:10.1016/j.dsx.2023.102744
31. Mihalache A, Popovic MM, Muni RH. Performance of an Artificial Intelligence Chatbot in Ophthalmic Knowledge Assessment. *JAMA Ophthalmol*. 2023;141(6):589. doi:10.1001/jamaophthalmol.2023.1144
32. Hopkins AM, Logan JM, Kichenadasse G, Sorich MJ. Artificial intelligence chatbots will revolutionize how cancer patients access information: ChatGPT represents a paradigm-shift. *JNCI Cancer Spectrum*. 2023;7(2):pkad010. doi:10.1093/jncics/pkad010
33. Huang J, Tan M. The role of ChatGPT in scientific communication: writing better scientific review articles. *Am J Cancer Res*. 2023;13(4):1148-1154.
34. Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a Radiology Board-style Examination: Insights into Current Strengths and Limitations. *Radiology*. 2023;307(5):e230582. doi:10.1148/radiol.230582
35. Singh S, Djalilian A, Ali MJ. ChatGPT and Ophthalmology: Exploring Its Potential with Discharge Summaries and Operative Notes. *Seminars in Ophthalmology*. 2023;38(5):503-507. doi:10.1080/08820538.2023.2209166
36. Xie Y, Seth I, Hunter-Smith DJ, Rozen WM, Ross R, Lee M. Aesthetic Surgery Advice and Counseling from Artificial Intelligence: A Rhinoplasty Consultation with ChatGPT. *Aesth Plast Surg*. 2023;47(5):1985-1993. doi:10.1007/s00266-023-03338-7
37. Gupta R, Park JB, Bisht C, et al. Expanding Cosmetic Plastic Surgery Research With

ChatGPT. *Aesthetic Surgery Journal*. 2023;43(8):930-937. doi:10.1093/asj/sjad069
38. Eggmann F, Weiger R, Zitzmann NU, Blatz MB. Implications of large language models such as ChatGPT for dental medicine. *J Esthet Restor Dent*. 2023;35(7):1098-1102. doi:10.1111/jerd.13046
39. Hoch CC, Wollenberg B, Lüers JC, et al. ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single-choice and multiple-choice board certification preparation questions. *Eur Arch Otorhinolaryngol*. 2023;280(9):4271-4278. doi:10.1007/s00405-023-08051-4
40. Humar P, Asaad M, Bengur FB, Nguyen V. ChatGPT Is Equivalent to First-Year Plastic Surgery Residents: Evaluation of ChatGPT on the Plastic Surgery In-Service Examination. *Aesthetic Surgery Journal*. 2023;43(12):NP1085-NP1089. doi:10.1093/asj/sjad130
41. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell*. 2023;6:1169595. doi:10.3389/frai.2023.1169595
42. Oh N, Choi GS, Lee WY. ChatGPT goes to the operating room: evaluating GPT-4 performance and its potential in surgical education and training in the era of large language models. *Ann Surg Treat Res*. 2023;104(5):269. doi:10.4174/astr.2023.104.5.269
43. Lund BD, Wang T, Mannuru NR, Nie B, Shimray S, Wang Z. ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Asso for Info Science & Tech*. 2023;74(5):570-581. doi:10.1002/asi.24750
44. Lee H. The rise of ChatGPT : Exploring its potential in medical education. *Anatomical Sciences Ed*. 2024;17(5):926-931. doi:10.1002/ase.2270
45. Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. *J Educ Eval Health Prof*. 2023;20:1. doi:10.3352/jeehp.2023.20.1
46. Lee JY. Can an artificial intelligence chatbot be the author of a scholarly article? *J Educ Eval Health Prof*. 2023;20:6. doi:10.3352/jeehp.2023.20.6
47. Adiguzel T, Kaya MH, Cansu FK. Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *CONT ED TECHNOLOGY*. 2023;15(3):ep429. doi:10.30935/cedtech/13152
48. Grassini S. Shaping the Future of Education: Exploring the Potential and Consequences of AI and ChatGPT in Educational Settings. *Education Sciences*. 2023;13(7):692. doi:10.3390/educsci13070692
49. Thurzo A, Strunga M, Urban R, Surovková J, Afrashtehfar KI. Impact of Artificial Intelligence on Dental Education: A Review and Guide for Curriculum Update. *Education Sciences*. 2023;13(2):150. doi:10.3390/educsci13020150
50. Halaweh M. ChatGPT in education: Strategies for responsible implementation. *CONT ED TECHNOLOGY*. 2023;15(2):ep421. doi:10.30935/cedtech/13036
51. Barrot JS. Using ChatGPT for second language writing: Pitfalls and potentials. *Assessing Writing*. 2023;57:100745. doi:10.1016/j.asw.2023.100745
52. Su Y, Lin Y, Lai C. Collaborating with ChatGPT in argumentative writing classrooms. *Assessing Writing*. 2023;57:100752. doi:10.1016/j.asw.2023.100752
53. Perkins M. Academic Integrity considerations of AI Large Language Models in the post-pandemic era: ChatGPT and beyond. *JUTLP*. 2023;20(2). doi:10.53761/1.20.02.07
54. Fergus S, Botha M, Ostovar M. Evaluating Academic Answers Generated Using ChatGPT. *J Chem Educ*. 2023;100(4):1672-1675. doi:10.1021/acs.jchemed.3c00087
55. Farrokhnia M, Banihashem SK, Noroozi O, Wals A. A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International*. 2024;61(3):460-474. doi:10.1080/14703297.2023.2195846
56. Emenike ME, Emenike BU. Was This Title Generated by ChatGPT? Considerations for

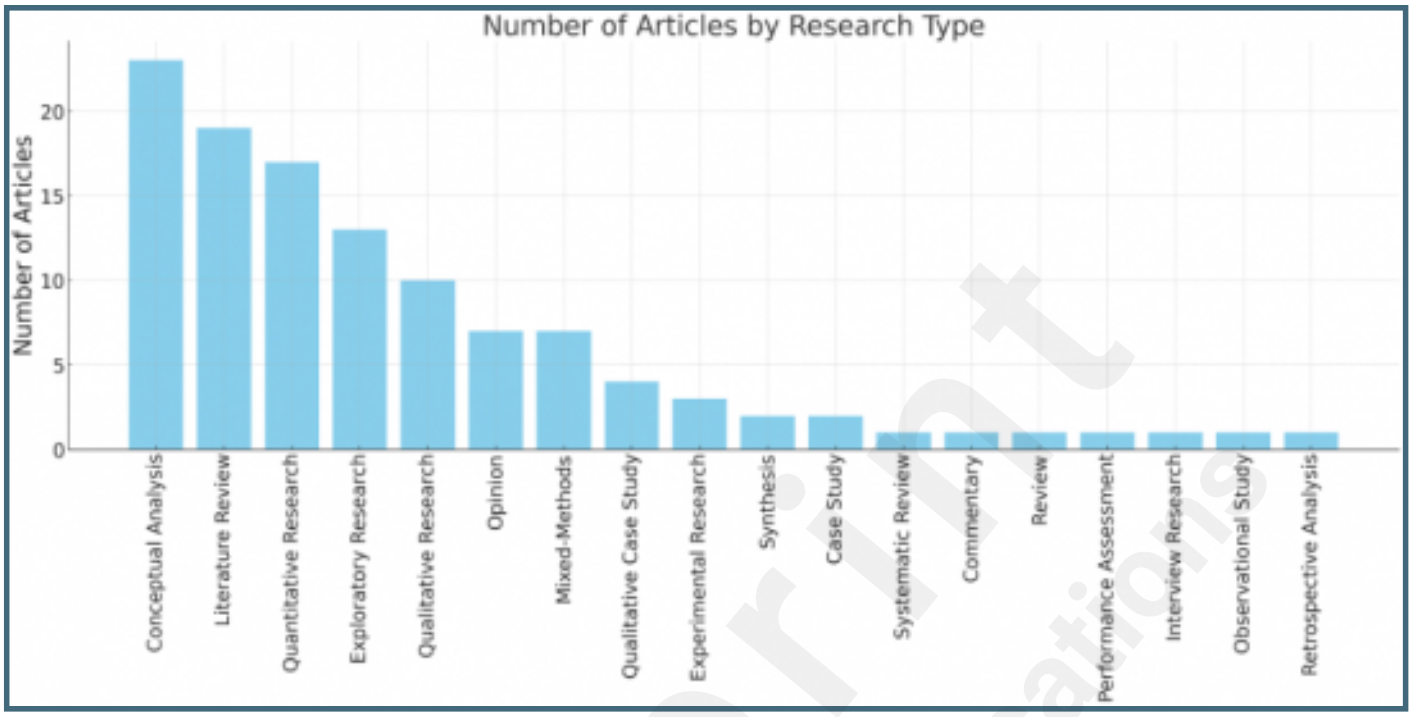
Artificial Intelligence Text-Generation Software Programs for Chemists and Chemistry Educators. *J Chem Educ*. 2023;100(4):1413-1418. doi:10.1021/acs.jchemed.3c00063
57. Cotton DRE, Cotton PA, Shipway JR. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*. 2024;61(2):228-239. doi:10.1080/14703297.2023.2190148
58. Cooper G. Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *J Sci Educ Technol*. 2023;32(3):444-452. doi:10.1007/s10956-023-10039-y
59. Takagi S, Watari T, Erabi A, Sakaguchi K. Performance of GPT-3.5 and GPT-4 on the Japanese Medical Licensing Examination: Comparison Study. *JMIR Med Educ*. 2023;9:e48002. doi:10.2196/48002
60. Crawford J, Cowling M, Allen KA. Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *JUTLP*. 2023;20(3). doi:10.53761/1.20.3.02
61. García-Peñalvo FJ. La percepción de la Inteligencia Artificial en contextos educativos tras el lanzamiento de ChatGPT: disrupción o pánico. *Educ Knowl Soc*. 2023;24:e31279. doi:10.14201/eks.31279
62. Lo CK. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*. 2023;13(4):410. doi:10.3390/educsci13040410
63. Tlili A, Shehata B, Adarkwah MA, et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learn Environ*. 2023;10(1):15. doi:10.1186/s40561-023-00237-x
64. Jeon J, Lee S. Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Educ Inf Technol*. 2023;28(12):15873-15892. doi:10.1007/s10639-023-11834-1
65. Strzelecki A. To use or not to use ChatGPT in higher education? A study of students' acceptance and use of technology. *Interactive Learning Environments*. Published online May 8, 2023:1-14. doi:10.1080/10494820.2023.2209881
66. Yan D. Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation. *Educ Inf Technol*. 2023;28(11):13943-13967. doi:10.1007/s10639-023-11742-4
67. Rahman MdM, Watanobe Y. ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Applied Sciences*. 2023;13(9):5783. doi:10.3390/app13095783
68. Hosseini M, Horbach SPJM. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Res Integr Peer Rev*. 2023;8(1):4. doi:10.1186/s41073-023-00133-5
69. Kohnke L, Moorhouse BL, Zou D. ChatGPT for Language Teaching and Learning. *RELC Journal*. 2023;54(2):537-550. doi:10.1177/00336882231162868
70. Sun GH, Hoelscher SH. The ChatGPT Storm and What Faculty Can Do. *Nurse Educ*. 2023;48(3):119-124. doi:10.1097/NNE.0000000000001390
71. Peres R, Schreier M, Schweidel D, Sorescu A. On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice. *International Journal of Research in Marketing*. 2023;40(2):269-275. doi:10.1016/j.ijresmar.2023.03.001
72. Dwivedi YK, Kshetri N, Hughes L, et al. Opinion Paper: "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*. 2023;71:102642. doi:10.1016/j.ijinfomgt.2023.102642
73. Roumeliotis KI, Tselikas ND. ChatGPT and Open-AI Models: A Preliminary Review. *Future Internet*. 2023;15(6):192. doi:10.3390/fi15060192
74. Anders BA. Is using ChatGPT cheating, plagiarism, both, neither, or forward thinking? *Patterns*. 2023;4(3):100694. doi:10.1016/j.patter.2023.100694
75. Du H, Teng S, Chen H, et al. Chat With ChatGPT on Intelligent Vehicles: An IEEE TIV

Perspective. *IEEE Trans Intell Veh*. 2023;8(3):2020-2026. doi:10.1109/TIV.2023.3253281
76. Chatterjee J, Dethlefs N. This new conversational AI model can be your friend, philosopher, and guide ... and even your worst enemy. *Patterns*. 2023;4(1):100676. doi:10.1016/j.patter.2022.100676
77. Kocoń J, Cichecki I, Kaszyca O, et al. ChatGPT: Jack of all trades, master of none. *Information Fusion*. 2023;99:101861. doi:10.1016/j.inffus.2023.101861
78. Vaithilingam P, Zhang T, Glassman EL. Expectation vs. Experience: Evaluating the Usability of Code Generation Tools Powered by Large Language Models. In: *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM; 2022:1-7. doi:10.1145/3491101.3519665
79. Hassani H, Silva ES. The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. *BDCC*. 2023;7(2):62. doi:10.3390/bdcc7020062
80. Taecharungroj V. “What Can ChatGPT Do?” Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. *BDCC*. 2023;7(1):35. doi:10.3390/bdcc7010035
81. Wu T, Terry M, Cai CJ. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In: *CHI Conference on Human Factors in Computing Systems*. ACM; 2022:1-22. doi:10.1145/3491102.3517582
82. Noy S, Zhang W. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*. 2023;381(6654):187-192. doi:10.1126/science.adh2586
83. Biswas SS. Role of Chat GPT in Public Health. *Ann Biomed Eng*. 2023;51(5):868-869. doi:10.1007/s10439-023-03172-7
84. Paul J, Ueno A, Dennis C. ChatGPT and consumers: Benefits, Pitfalls and Future Research Agenda. *Int J Consumer Studies*. 2023;47(4):1213-1225. doi:10.1111/ijcs.12928
85. Rana MS, Nobi MN, Murali B, Sung AH. Deepfake Detection: A Systematic Literature Review. *IEEE Access*. 2022;10:25494-25513. doi:10.1109/ACCESS.2022.3154404
86. Holzinger A, Keiblinger K, Holub P, Zatloukal K, Müller H. AI for life: Trends in artificial intelligence for biotechnology. *New Biotechnology*. 2023;74:16-24. doi:10.1016/j.nbt.2023.02.001
87. Korzynski P, Mazurek G, Altmann A, et al. Generative artificial intelligence as a new context for management theories: analysis of ChatGPT. *CEMJ*. 2023;31(1):3-13. doi:10.1108/CEMJ-02-2023-0091
88. Currie GM. Academic integrity and artificial intelligence: is ChatGPT hype, hero or heresy? *Seminars in Nuclear Medicine*. 2023;53(5):719-730. doi:10.1053/j.semnuclmed.2023.04.008

Supplementary Files

Figures

This is a bar graph that contains the number of articles counted in the study and their categorizations. This figure is also embedded in the manuscript.



Multimedia Appendixes

Contains data collected and links to specific ChatGPT threads used to perform the study. Includes items from 8 appendices.
URL: http://asset.jmir.pub/assets/fe610bd584f1122c3b9abf1753820dec.docx