Let ChatGPT Write Your Master's Thesis: An Exploratory Case-Study

Pål Joranger, Sara Rivenes Lafontan, Asgeir Brevik

Submitted to: JMIR Formative Research
on: February 28, 2025

Table of Contents

Let ChatGPT Write Your Master's Thesis: An Exploratory Case-Study

Pål Joranger[1] PhD; Sara Rivenes Lafontan[1] PhD; Asgeir Brevik[1] PhD

[1]Department of Nursing and Health Promotion Faculty of Health Sciences OsloMet – Oslo Metropolitan University Oslo NO

Corresponding Author:
Asgeir Brevik PhD
Department of Nursing and Health Promotion
Faculty of Health Sciences
OsloMet – Oslo Metropolitan University
P.O. Box 4 St. Olavs plass
Oslo
NO

Abstract

Background: Large language models (LLMs) can aid students in mastering a new topic fast but for the educational institutions responsible for assessing and grading the academic level of students it can be difficult to discern whether a text originates from a student's own cognition or if it is synthesized by an LLM. Universities have traditionally relied on a submitted written thesis as proof of higher-level learning, on which to grant grades and diplomas. Ubiquitous availability of LLMs challenges this practice.

Objective: In this study we assumed the role of hypothetical health science master's student actors looking to leverage the full power of an LLM in completing scientific research paper manuscripts that could be submitted for master's thesis graduation.

Methods: In an exploratory case-study we used ChatGPT to generate two research papers as conceivable student submissions for master's thesis graduation from a health science master's program. One paper simulated a qualitative and another simulated a quantitative research project.

Results: Using a stepwise approach we prompted ChatGPT to 1) synthesize two credible datasets, and 2) generate two manuscripts, in less than a day, that—in our judgment—would have been able to pass as credible graduation research papers at the health science master's program the authors are currently affiliated with.

Conclusions: Our demonstration highlights the ease with which an LLM can synthesize research data, conduct scientific analyses, and produce credible research papers for a master's graduation. To uphold the integrity of academic standards, we recommend master's programs to prioritize oral examinations and school exams. This shift is crucial to ensure a fair and rigorous assessment of higher-order learning and abilities at the master's level.

(JMIR Preprints 28/02/2025:73248)
DOI: https://doi.org/10.2196/preprints.73248

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
No. Please do not make my accepted manuscript PDF available to anyone.

Original Manuscript

Let ChatGPT Write Your Master's Thesis: An Exploratory Case-Study

Pål Joranger[1], Sara Rivenes Lafontan[1] and Asgeir Brevik[1]*
[1]Affiliation; Institute of Nursing and Health Promotion, Oslo Metropolitan University, Oslo, Norway.
*Correspondence: asgeir@oslomet.no

Abstract

Background:
Large language models (LLMs) can aid students in mastering a new topic fast but for the educational institutions responsible for assessing and grading the academic level of students it can be difficult to discern whether a text originates from a student's own cognition or if it is synthesized by an LLM. Universities have traditionally relied on a submitted written thesis as proof of higher-level learning, on which to grant grades and diplomas. Ubiquitous availability of LLMs challenges this practice.

Objective:
In this study we assumed the role of hypothetical health science master's student actors looking to leverage the full power of an LLM in completing scientific research paper manuscripts that could be submitted for master's thesis graduation.

Methods:
In an exploratory case-study we used ChatGPT to generate two research papers as conceivable student submissions for master's thesis graduation from a health science master's program. One paper simulated a qualitative and another simulated a quantitative research project.

Results:
Using a stepwise approach we prompted ChatGPT to 1) synthesize two credible datasets, and 2) generate two manuscripts, in less than a day, that—in our judgment—would have been able to pass as credible graduation research papers at the health science master's program the authors are currently affiliated with.

Conclusions:
Based on our demonstration of how easy it is to harness the power of an LLM to synthesize research data; perform scientific analyses and write credible research papers for a master's graduation, we recommend health master's programs to put more emphasis on oral examination and school exams to ensure fair judgement of higher order learning and abilities at the master's level.

Keywords: master's thesis; large language models; LLMs; ChatGPT; LLMs; health science; qualitative; quantitative

Introduction

Students represent a group that benefits from the use of LLMs. By Using LLMs, students can gain knowledge on virtually any topic as well as edit or create scholarly text instantaneously [1–3]. In a survey from 2023 it was documented that 89% of colleague students in America used ChatGPT to complete their assignments and 53% used the tool for writing papers [4].

Although potentially beneficial for learning, the ubiquitous availability and use of LLMs challenge traditional methods of measuring the quality of student's academic performance. To obtain a master's degree, a written thesis is traditionally required. If theses, independent student research projects or written home exams now can be easily completed using LLMs, their role as a robust demonstration of higher-level skills and learning might not be justified going forward.

Despite their powerful capabilities, LLM's are still relatively new tools with major improvements yet to emerge making them even more powerful in the future, autonomous executing AI agents being an example [3,5]. It may be argued that many corners of academia are already struggling to keep up with the pace of AI development [6]. While plagiarism detection tools are widely used at universities, it has been observed that traditional plagiarism detection tools may struggle to detect AI-generated text [7]. As lecturers and teachers grading exams at the university, there is a need to assess how much of a typical master's thesis can now be written by LLMs? It also raises concerns that faculty staff could be left with the unappealing task of grading a large amount of almost similar student projects written largely by LLMs, falsely believing that they are assessing the student's academic ability. There is an urgent need for discussion on how to retain the value of university diplomas in an academic world rattled by the widespread adoption of LLMs.

As health science lecturers and master thesis censors at a Norwegian university, we were curious to test the ability of one of the most accessible LLMs to assist budding master thesis authors in completing their graduation thesis. At the university the authors are affiliated with, master's students can choose to submit their graduation work in the form of a monography, or as a scientific article preceded by a short introduction. Hence, we set out to briefly test the power of ChatGPT to assist in the professional analysis and presentation of health science data—both in the realm of qualitative and quantitative data analysis as well as the writing of an academic paper based on the results. We evaluated whether the quality of LLM generated master thesis research articles would be sufficient to pass the final exam at the master's level at the master's program we are currently affiliated with. We end by discussing implications for teachers/evaluators and the prospects of assessing learning outcome achievement at the master's level, in a new LLM disrupted academic era.

Materials and Methods

In an exploratory case study, we asked ChatGPT to synthesize two datasets: one qualitative and one quantitative, inspired by two existing real-world datasets the authors had previously analyzed. The generation of the synthetic datasets and subsequent data analysis took place in October and November of 2024.

Generation of qualitative data

We started by outlining the study aim and instructed ChatGTP4o mini to prepare interview transcripts of seven students with different levels of education, age and gender. We also prompted ChatGPT to develop an interview guide (Supplementary material 1.1). The first transcripts generated

by ChatGTP4o mini were quite short hence in a follow-up prompt to ensure that each interview reflected 30-45 minutes of conversation, we request-ed interview transcripts that included longer and more detailed responses detailing specific events relevant to the study aim, including respondents' experiences and feelings. We also prompted for the artificial interviewer to ask follow-up questions, emulating a natural conversation. We prompted for one interview transcript at a time to ensure that these aspects were included in each transcript. This resulted in seven transcripts that were considered suitable for further analysis. The qualitative dataset is shown in Supplementary material 1.2.

Generation of quantitative dataset

For the quantitative data, we used ChatGPTo1 to develop a script that was run in ChatGPT4o to produce the data matrix. To work with meaningful associations that resembled what could be found in international literature, we had to provide ChatGPT with information about how our synthetic dataset's 10 variables correlated with each other in pairs. To avoid inconsistencies in how the variables correlated as an overall system, we relied on a correlation matrix from real data for similar variables. All the correlation coefficients given to ChatGPT were discretionarily slightly changed (usually increased slightly) before admission, but without changing the direction (positive or negative) of the correlation. The quantitative dataset is shown in Supplementary material 2.1.

Overall approach to ChatGPT assisted generation of manuscripts

Conceptually our workflow advanced in the following logical sequence (with continuous prompt optimization based on the output from the LLM): a) upload dataset, prompt the LLM to analyze and present results using specified method; b) suggest methods chapter based on user pre-determined context; c) suggest introduction chapter, with emphasis on providing relevant scientific citations; d) develop a discussion, with emphasis on integrating relevant scientific citations; e) suggest abstract and title; f) combine elements to a final manuscript, update list of references. We opted for a "bite-sized" chapter by chapter approach to our ChatGPT writing, because in our experience, ChatGPT tends to respond to single prompts with a relatively limited amount of text output.

Preparation of qualitative manuscript

To conduct the data analysis and generate the research article we prompted ChatGTP4o mini to develop a qualitative analysis of the seven transcripts, repeating the study aim and areas of interest. We also specified the method of qualitative data analysis that we wanted to use, thematic analysis by Braun and Clarke, and that 3-4 themes should be developed with two illustrative statements for each theme [8]. Following ChatGTP4o's qualitative data analysis output, we asked it to develop the results chapter for a manuscript for a scientific publication. The initial text output lacked sufficient detail, and an additional prompt was needed to request a more elaborate analysis. ChatGPT was then prompted to write introduction, methods chapter, discussion, methodological discussion and conclusion as well as the abstract. ChatGPT was also prompt-ed to suggest suitable titles to the manuscript and state which one it recommended. After prompting to move all the citations to the bottom of the text, we made a Microsoft Word text file of the complete manuscript. The file was then uploaded to ChatGPT with the prompt to suggest areas of improvement. A final set of prompts was necessary for the in-corporation of the suggested improvements for each chapter of the manuscript. (See Supplementary material 1.3 for full list of prompts for development of qualitative manuscript).

Preparation of quantitative manuscript

We started by informing ChatGPT4o about the overarching research question. Having specified which variables were dependent variables, explanatory variables and control variables, a multiple

linear regression analysis was requested. We explained which parameters we wanted to have estimated and included in a table. Next, we prompted ChatGPT4o to do a Spearman's rank-correlation analysis where all the variables used above were correlated against each other two by two with corresponding r and p-values. Finally, a descriptive analysis was requested. For all three analyses, we prompted ChatGPT to set up the results in tables suitable for publication in scientific journals.

ChatGPT was prompted to prepare the results chapter that reports the results of the three analyses with an emphasis on the multiple regression analysis. For the preparation of this chapter and for the subsequent ones, we instructed ChatGPT to use the STROBE checklist [9] for the preparing cross-sectional analyses. The checklist was up-loaded as an attachment to the prompt. In separate dialogs, we asked ChatGPT to prepare the methods chapter and the introductory chapter, respectively. We promoted ChatGPT to work on the discussion in two separate segments, one discussing the methods and one discussing the results. Conclusions were requested at the end.

We entered all the sub-bibliographies from individual chapters into a document and prompted ChatGPT4o to remove duplicates, check for authenticity, and to print the list in APA style. In separate prompts we requested ChatGPT4o to prepare a summary and to suggest a title for the article. The various parts were combined manually into one manuscript. Lastly, ChatGPT4o was prompted to suggest improvements to its own manuscript. Based on ChatGPT4o's suggestions for improvements, we instructed ChatGPT4o to refine the manuscript, chapter by chapter, and in the end the summary. The various optimized parts were edited together manually into a complete manuscript. All results from the bi- and multivariate analyses performed by ChatGPT were tested against analyses in SPSS 28 and StataMP 18. The full list of prompts used for the quantitative manuscript is shown in Supplementary material 2.2.

Results

Our results consist of two full length scientific paper manuscripts (see Supplementary material 1.4 and 2.3 for the qualitative and quantitative manuscript respectively). In the following we present our observations of the strengths and weaknesses of the manuscripts developed for our particular use, i.e. to emulate a health science master students research article submitted for graduation, followed by a short presentation of our judgement of the final full-size manuscripts, as well as estimates of total time spent for dataset and manuscript development.

Examination and assessment of the qualitative manuscript

The qualitative manuscript produced contained the key elements expected of a scientific article (Supplementary material 1.4). It was concise, with 4529 words. The thematic analysis reflected the study aim, with a clear identification of themes and relevant illustrative quotes included for each theme. The generated manuscript provided a logical basis for discussion and the abstract followed standard conventions, offering an adequate summary of the main findings and included reference to Braun and Clarke’s six-phase analytical method. The manuscript presented four main themes—motivation to mentor, personal and professional growth, challenges faced, and the need for enhanced training. However, the manuscript had some weaknesses. The thematic analysis lacked depth in exploring the various experiences of the “participants” and many interpretations were superficial. For example, the theme “Motivation to become a mentor” highlights the participants desire to support others and develop leadership skills while largely reiterating generic motivations

without exploring individual nuances or cultural influences on these motivations. There were also repetition and overlap in certain quotes. While there were some references to existing literature, the use of references was limited and as such the manuscript failed to connect findings to in broader evidence based scientific discussion. In addition, out of 11 references, 7 were suspected to have been fabricated by ChatGPT. Despite the weaknesses, the manuscript was assessed to have received a pass grade according to the university's censor guidelines. That said, it is doubtful it would have passed if the fabricated references were discovered by the censors.

The generation of the manuscript itself was both fast and efficient. While the total time spent generating both the synthesized data and the manuscript was approximately 2,5 hours, approx. 1,5 hours were used to generate the synthesized data as it required several prompts to ensure detailed personal reflections and feelings.

Examination and assessment of the quantitative manuscript

The resulting manuscript (Supplementary material 2.3) includes the main elements one would expect and generally aligns with the STROBE guidelines. It is relatively concise, with 2930 words excluding the abstract, tables, and references. ChatGPT performed correlation analyses and multiple linear regression analyses, but the multiple linear regression did not include robust standard error adjustments. Chapter coherence and focus on the research question is deemed satisfactory. Relevant discussions on methodological challenges and comparisons with existing literature are present and seemed genuine, while the existence of one publication could not be confirmed[1]. The abstract followed the standard thematic structure and provides a satisfactory summary.

One area where ChatGPT failed was in conducting multiple regression analyses based on robust standard errors. After the initial multiple regression analyses were per-formed, we asked ChatGPT to check if the assumptions for these analyses were met. According to ChatGPT, they were met except for potential issues with the assumption of homoscedasticity. When we asked ChatGPT how this issue could be resolved, it provided six different methods to address the problem. We chose the first method it suggested, which was to conduct multiple regression analyses based on robust standard errors. We then prompted ChatGPT to perform these analyses, and the transformer produced new estimates for the relevant parameters, explicitly stating, "Here are the results from the multiple regression analysis with robust standard errors summarized in a table." The problem was that this table was identical to the original one.

Further miscalculations by ChatGPT were exposed by the fact that 9 out of the 28 bivariate correlation analyses (Spearman) were incorrect. This can be observed by comparing Table 3 with the Spearman correlation matrix in Supplementary material 2.4, estimated using SPSS. All bivariate analyses between life satisfaction and the other variables appear to have been correctly calculated by ChatGPT4o. We checked whether the 9 discrepancies aligned with the corresponding bivariate correlation analyses performed using Pearson rather than Spearman correlation, but this was not the case.

Assuming the examiners do not conduct their own analyses and therefore do not detect the errors in the "robust" analyses and the miscalculations of Spearman correlations, we assess that this manuscript would pass as part of a master's thesis. At the university we are affiliated with, it is not common practice for examiners to perform their own control analyses based on students' data. However, if the examiners did indeed do their own calculations, and discovered the error, we believe the fictional student could still pass albeit with a lower grade. Readers can make their own assessments of the quality of the manuscript, which is attached as Supplementary material 2.3.

[1]Gonzalez, A., Luces, R., & Popkin, B. (2018). Cigarette smoking and mental health: A longitudinal perspective. Tobacco Control, 27(4), 442–448.

A total of approximately 3.5 hours was spent preparing the manuscript, including verifying the authenticity of the references. This excludes the time spent creating the synthetic data and the time required to prepare the table of variables used. The simulated situation assumed that the student was using secondary data, with an accompanying de-scription of included variables. We excluded the time spent verifying calculations done in SPSS and Stata, as we assumed that the students relied solely on the analyses performed by ChatGPT.

Discussion

This case study demonstrates that it is quite easy to develop a scientific manuscript to be submitted as part of a master thesis using the ChatGPT. While the ChatGPT generated manuscripts might not represent a well performing student's highest potential, we rate them sufficient to score a pass grade at the health science master's program the authors are currently affiliated with. Further work with prompt optimization could potentially have improved our manuscript examples. We also demonstrated that it is possible to use ChatGPT to synthesize credible research data.

It was surprisingly easy to generate data, in particular qualitative data. With a couple of additional prompts, we were also able to generate a quantitative dataset. This synthetic quantitative dataset appears to possess sufficient complexity and internal consistency to be useful as an open dataset for illustrating our experiments with ChatGPT. However, readers who conduct a closer examination of the data may observe that several variables exhibit distributions that are not entirely realistic. That fact that ChatGPT can be used to falsify or fabricate data, and the associated potential for academic misconduct, has been recognized by other authors [10,11].

It could be argued that collectively the authors possess more research experience than a typical master's student, and as such are better equipped to compose effective prompts for the fabrication of research data and master's thesis elements. Clearly, an average student starting from scratch, with limited experience, would likely take longer to arrive at useful results. At the same time, one would expect that students representing a generation well versed in online information retrieval would be able to quickly acquire the know-how of generating a master's thesis using LLMs. One can easily envision a predefined, standardized set of LLM prompts designed for the efficient generation of master's theses. In an evolving "arms race," emerging AI agents could further enhance students' ability to circumvent university anti-plagiarism systems.

ChatGPT4 produced shorter text than expected. This tendency of ChatGPT4 to produce well written superficial answers that fail to capture the level of factual detail, terminology and nuance that is often expected in academia has been observed by other authors [12]. By issuing more specific prompts, with clear expectations for context and output, it was usually possible to extract a lot more detail than what was generated from the initial prompt(s). Prompt engineering for academic writers has been covered by Giray [13]

ChatGPT sometimes generated false references when instructed to provide references supporting its claims. This tendency to, despite prompts, omit scientific citations and generate non-existing references has been reported by other authors [11,12,14]. ChatGPT's suggested references could be compromised in two ways: 1) they could be non-existent, and 2) they could be real but not relevant or at least not the ones that would have been chosen by the human expert in the specific field. While it is easy to check for the existence of cited references (1) checking for the quality and relevance of the cited references (2) requires a modicum of expertise.

As LLMs frequently generate non-existing citations and incorrect statements authors have warned against uncritical use of ChatGPT in scientific writing [11,14]. Indeed, in-depth knowledge of the subject matter at hand, including the methods employed, would represent an advantage for optimal

utilization of LLMs.

In the ChatGPT assisted quantitative manuscript we manually checked the accuracy of the statistical calculations performed by ChatGPT. Although the mistakes did not stand out as obvious and easily detectable, we did, on closer scrutiny, discover some errors and irregularities. The errors were not grave enough to set off an immediate red flag and would likely go undetected should the censors not perform their own control calculations. Nevertheless, our experience illustrates that research integrity may suffer from over-reliance on LLMs in their current state.

Due to budgetary constraints and increased student volume, master's thesis censors are not necessarily subject matter experts, often unable to verify each citation of the many master's theses they are required to evaluate. Hence, suboptimal references could in many instances be expected to slip through unnoticed and fail to draw the critical remarks they deserved in an exam situation.

One argument that is made against the prospects of am LLM-assisted streamlined master's thesis expressway is that the student's supervisor would surely detect any attempts at cheating. This belief may be based on how things used to be in a bygone era. In a modern world where technological tools increasingly allow for economically efficient online based master programs, daily contact between faculty staff and graduate students is no longer guaranteed. With increasingly limited oversight, it might be tempting for the graduate student to rely heavily on AI assistance to complete their academic tasks, especially when the universities anti-plagiarism tools are unable to detect AI generated text [15]. Furthermore: Should heavy reliance on ChatGPT be considered cheating at all or is it just a new power tool for all students to enjoy un-restricted, like they would be able to in a real-life work situation.

While the challenges related to student's use of LLMs by now are well-known in higher education, we have been unable to identify clear solutions from the current literature. Master's programs still rely heavily on thesis-based assessment; some exclusively rely on thesis-based assessment. Based on our case study, demonstrating how easy it is to cheat by employing heavy AI assistance in the development of a master's thesis, it is our belief that assessment of academic achievements should no longer be based purely on the student's written work. An oral examination is often conducted after the submitted master thesis to determine an overall grade for the thesis. Traditionally—at least at our local university—the written thesis has been weighed more than the student's performance at the oral defense. The term "adjusting oral defense" has been used in-house by faculty staff and the university administration, implying that only minor adjustments—like one grade up or down—is to be expected following oral examination. As a perhaps unfortunate but necessary consequence of the LLM revolution we suggest flipping the script on this practice. When it comes to grading the student's academic achievement and maturity, placing more emphasis on the results of the oral examination of the student than on the student's submitted thesis now makes sense. After all, if universities are not able to enforce correct, and what they consider to be the most ethical student use of LLMs, it might be argued that the reasonable thing to do is to lift most if not all restrictions, and instead rely primarily on oral examinations where students are given the opportunity to demonstrate deep knowledge of their subject, scientific reasoning, the relevant analytical steps of their analysis, as well as the inherent strengths and weaknesses of the methods used in their thesis. This switch does not have to incur extra expenses on an often-challenged university economy. Evaluator resources spent on repeated reading of students written texts could be better spent on broad oral examinations. Oral exams could include questions from the student's written master's thesis but also a set of thesis independent questions on scientific methods that required less preparation from the examiner.

One limitation with our ChatGPT generated master's manuscript assessment is that the finished

product was neither put through a proper peer-assessment nor submitted to the university examination system as a sham student project. As master program teachers, supervisors, course leaders and master thesis examiners we believe we can assess whether the fake manuscripts would have been able to obtain a pass grade. A master the-sis contains more than just the scientific manuscript that we have used as an example in our case study. However, we believe that the process of developing the manuscripts is relevant to illustrate how easy it is to produce scientific work at a high level using LLMs. We only made two example manuscripts. More manuscripts utilizing a diverse set of analyses could perhaps have revealed relevant nuances between analytical methods. Given the highly competitive field of AI tools, we could have tested more than one LLM. Further prompt optimization could have reduced the number of false references and prompting for the use of the COREQ checklist [16] or similar tool, could have further enhanced the quality of the qualitative manuscript.

Conclusions

While the universally available LLMs constitute a family of extraordinarily capable new software tools that promise to revolutionize scientific productivity, they will also be of great help for students at all levels. An emerging realization is that ubiquitous use of LLMs will pose a challenge for academic institutions entrusted with the task of ensuring a certain level of acquired knowledge by graduating students, and issuing diplomas based on demonstrated academic achievements. In the present study, we have demonstrated that LLMs can produce credible student master's thesis manuscripts at a fraction of the time normally allocated to the completion of a thesis submitted for graduation. We also demonstrate that LLMs can create entirely new artificial datasets with relative ease when modeled upon previously published research.

Traditionally, universities have relied heavily on a written academic thesis to evaluate academic achievement and issue grades. With the introduction of LLMs, capable of generating credible scientific texts, academic institutions now risk issuing grades based on machine generated output. Based on our role-playing experience as hypothetical busy student actors looking for a heavily AI-assisted fast-track to a master's thesis, we recommend against pure text-based ascertainment of academic achievement for graduate students. Increased emphasis on oral examinations for grade defining graduation work is a logical consequence of a looming LLM caused disruption of the ingrained methods of assessing students' performance in higher education.

Supplementary Materials: The following supporting information can be downloaded at: www.XXX.com, S1.1: Interview guide; S1.2: Prompts qualitative; S1.3: Qualitative dataset; S1.4: Qualitative manuscript; S2.1: Quantitative dataset; S2.2: Prompts quantitative; S2.3: Quantitative manuscript; S2.4: Spearman correlations.

Author Contributions: Conceptualization, PJ and AB; methodology, PJ, SRL, AB; investigation, PJ, SRL, AB; data curation, PJ, SRL, AB.; writing—original draft preparation, PJ, SRL, AB.; writing—review and editing, PJ, SRL, AB; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable for studies not involving humans or animals.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is available in supplementary files.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Meyer JG, Urbanowicz RJ, Martin PCN, O'Connor K, Li R, Peng P-C, Bright TJ, Tatonetti N, Won KJ, Gonzalez-Hernandez G, Moore JH. ChatGPT and large language models in academia: opportunities and challenges. BioData Min 2023 Jul 13;16(1):20. doi: 10.1186/s13040-023-00339-9

2. Patil R, Gudivada V. A Review of Current Trends, Techniques, and Challenges in Large Language Models (LLMs). Appl Sci Multidisciplinary Digital Publishing Institute; 2024 Mar 1;14(5):2074. doi: 10.3390/app14052074

3. Wang S, Xu T, Li H, Zhang C, Liang J, Tang J, Yu PS, Wen Q. Large Language Models for Education: A Survey and Outlook. arXiv; 2024. doi: 10.48550/arXiv.2403.18105

4. Yu H. Frontiers | Reflection on whether Chat GPT should be banned by academia from the perspective of education and teaching. Front Psychol 2023 Jun 1; doi: 10.3389/fpsyg.2023.1181712

5. Pachegowda C. The Global Impact of AI-Artificial Intelligence: Recent Advances and Future Directions, A Review. arXiv; 2023. doi: 10.48550/arXiv.2401.12223

6. Strzelecki A. 'As of my last knowledge update': How is content generated by ChatGPT infiltrating scientific papers published in premier journals? Learn Publ 2025;38(1):e1650. doi: 10.1002/leap.1650

7. Santra PP, Majhi D. Scholarly Communication and Machine-Generated Text: Is it Finally AI vs AI in Plagiarism Detection? J Inf Knowl 2023 Jun 30;175–183. doi: 10.17821/srels/2023/v60i3/171028

8. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol Routledge; 2006 Jan;3(2):77–101. doi: 10.1191/1478088706qp063oa

9. von Elm E, Altman D, Egger M, Pocock S, Gøtzsche P, Vandenbroucke J. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PubMed 2007; Available from: https://pubmed.ncbi.nlm.nih.gov/17941714/ [accessed Jan 16, 2025]

10. Taloni A, Scorcia V, Giannaccare G. Large Language Model Advanced Data Analysis Abuse to Create a Fake Data Set in Medical Research. PubMed 2023; Available from: https://pubmed.ncbi.nlm.nih.gov/37943569/?dopt=Abstract [accessed Jan 10, 2025]

11. Hosseini M, Rasmussen LM, Resnik DB. Using AI to write scholarly publications. Account Res Taylor & Francis; 2024 Dec 6;31(7):715–723. PMID:36697395

12. Reis F, Lenz C, Gossen M, Volk H, Drzeniek N. Practical Applications of Large Language Models for Health Care Professionals and Scientists. PubMed 2024; Available from: https://pubmed.ncbi.nlm.nih.gov/39235317/ [accessed Dec 20, 2024]

13. Giray L. Prompt Engineering with ChatGPT: A Guide for Academic Writers. PubMed. Available from: https://pubmed.ncbi.nlm.nih.gov/37284994/ [accessed Jan 10, 2025]

14. Salvagno M, Taccone FS, Gerli AG. Artificial intelligence hallucinations. Crit Care 2023 May 10;27(1):180. doi: 10.1186/s13054-023-04473-y

15. Pudasaini S, Miralles-Pechuán L, Lillis D, Llorens Salvador M. Survey on AI-Generated Plagiarism Detection: The Impact of Large Language Models on Academic Integrity. J Acad Ethics Springer; 2024 Nov 4;1–34. doi: 10.1007/s10805-024-09576-x

16. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec 1;19(6):349–357. doi: 10.1093/intqhc/mzm042

Supplementary Files

Multimedia Appendixes

Supplementary_material_1_1_interview_guide.
URL: http://asset.jmir.pub/assets/ca50a64fd653cc3e412e7314e8548162.docx

Supplementary_material_1_2_prompts_qualitative.
URL: http://asset.jmir.pub/assets/0f2fba009e3c8316fe1b92784826bf46.docx

Supplementary_material_1_3_qualitative_dataset.
URL: http://asset.jmir.pub/assets/6334d6ce464ad7b70bba3b86588f1f97.docx

Supplementary_material_1_4_qualitative_manuscript.
URL: http://asset.jmir.pub/assets/f726a8d86049aa5b4ae8b68ba93a4a22.docx

Supplementary_material_2_1_quantitative_dataset.
URL: http://asset.jmir.pub/assets/b4fb262f88c2071f6f32a953bb262057.xlsx

Supplementary_material_2_2_prompts_quantitative.
URL: http://asset.jmir.pub/assets/534db2d574fc2a530eecfe977e85c75e.docx

Supplementary_material_2_3_quantitative_manuscript.
URL: http://asset.jmir.pub/assets/d4fd73d81d819679f16f115b3aa99423.docx

Supplementary_material_2_4_spearman_correlations.
URL: http://asset.jmir.pub/assets/1bae70709247d80057e3cd74d44072be.docx