Automatic Reviewing of Conference Papers in Healthcare and Other Sciences Using ChatGPT

Matjaž Gams
Department of Intelligent Systems
Jozef Stefan Institute
Ljubljana Slovenia
matjaz.gams@ijs.si

Luka Bolješić
University of Ljubljana
Faculty of Computer and Information Science
Ljubljana Slovenia
luka.boljesic0@gmail.com

Alex Korlat
University of Ljubljana
Faculty of Computer and Information Science
Ljubljana Slovenia
alex.korlat.ak@gmail.com

Primož Kocuvan
Department of Intelligent Systems
Jozef Stefan Institute
Ljubljana Slovenia
primoz.kocuvan@ijs.si

ABSTRACT

The emergence of artificial intelligence (AI) has significantly impacted various fields, including the peer review process in academic and scientific research. ChatGPT, a large language model developed by OpenAI, has shown potential in automating and enhancing the review of conference healthcare papers. Our HealthReview AI tool can process and analyze large volumes of text rapidly, providing feedback and insights that streamline the peer review process, reduce human workload, and increase efficiency. This paper presents a web application developed using the Flask framework that enables users to upload PDF files containing research papers and utilizes ChatGPT to generate reviews for each paper. The methodology, results and potential implications of this application are discussed, highlighting both the advantages and the challenges of integrating AI into the academic review process.[1]

KEYWORDS

Artificial Intelligence, ChatGPT, Peer Review, Healthcare, Scientific Research, Flask Framework, PDF Processing, Academic Writing, Conference Papers

POVZETEK

Pojav umetne inteligence (UI) je pomembno vplival na različna področja, vključno s postopkom strokovnega pregleda v akademskih in znanstvenih raziskavah. ChatGPT, velik jezikovni model, ki ga je razvil OpenAI, je pokazal potencial za avtomatizacijo in izboljšanje pregleda medicinskih konferenčnih prispevkov. To orodje UI lahko hitro obdela in analizira velike količine besedil ter zagotovi povratne informacije in vpoglede, ki poenostavijo postopek strokovnega pregleda, zmanjšajo delovno obremenitev in povečajo učinkovitost. Ta članek predstavlja spletno aplikacijo HealthReview, razvito s pomočjo ogrodja Flask, ki uporabnikom omogoča nalaganje datotek PDF, ki vsebujejo raziskovalne prispevke, in uporablja ChatGPT za generiranje pregledov za vsak prispevek. Obravnavane so metodologija, rezultati in možni vplivi te aplikacije, ki poudarjajo tako prednosti kot izzive integracije UI v akademski pregledni postopek.

KLJUČNE BESEDE

Umetna inteligenca, ChatGPT, strokovni pregled, zdravstvo, znanstvene raziskave, Flask ogrodje, obdelava PDF, akademsko pisanje, konferenčni prispevki

1 Introduction

The integration of artificial intelligence (AI) across various domains is significantly transforming complex processes, including the peer review of conference papers in healthcare and other sciences. Among AI tools, ChatGPT, developed by OpenAI, stands out for its potential to automate and enhance the review process. Its ability to quickly analyze large volumes of text and provide insightful feedback could streamline peer reviews, reduce human workload, and improve overall efficiency.

Recent studies highlight the diverse applications of ChatGPT in healthcare education, research, and practice. For example, it has been shown to improve scientific writing, analyze datasets, and aid drug discovery [1, 2, 3]. Additionally, its role in generating paraphrased content and literature reviews indicates the potential to expedite academic tasks, although concerns about originality and accuracy persist [4, 5].

ChatGPT's integration into medical literature reviews has been explored, demonstrating its ability to synthesize medical knowledge, though ethical and accuracy issues require further research [6, 7]. Beyond healthcare, ChatGPT enhances research efficiency across various scientific fields. It effectively generates Boolean queries for systematic reviews and supports rapid

Information Society 2024, 7–11 October 2024, Ljubljana, Slovenia

literature searches [8]. The AI's potential to streamline peer reviews and address biases, is also evident, though managing issues like bias, plagiarism, and inaccuracies remain crucial to maintaining academic integrity [9, 10, 11].

In our opinion, tools like ChatGPT offer significant opportunities to enhance the peer review process. However, careful deployment is necessary to ensure ethical considerations, accuracy, and the preservation of academic integrity. This paper explores these aspects and presents insights into effectively integrating ChatGPT into the peer review process designed with our system HealthReview.

The paper consists of Section 1 Introduction, section 2 Methodology where the system is presented. Results are demonstrated in Section 3, and the paper concludes with a discussion.

2 Methodology

When tested, GPT-4o was already able to reply to all basic questions, e.g. “Is grammar in that text correct”. Therefore, the first task was to create a list of commands to perform the sequence of the review, chosen for the Information Society conference (is.ijs.si). The second task was to fine-tune the process since several output issues were not as desired. The third task was to include additional knowledge, and that was executed by including the Insieme knowledge base as the core medical information. In this way, the HealthReview performance achieved reasonable levels.

The Insieme platform was selected for integration with GPT (https://www.youtube.com/watch?v=ozAporFOw64). This platform features a user-friendly interface that enables users to efficiently access valuable healthcare information from a centralized website, either via manual search or through the Insieme search function. Insieme serves as the successor to the national Electronic and Mobile Health (eHealth) initiative, a project that was characterized by collaboration among 15 partners. Furthermore, the platform's development has been significantly shaped by insights derived from the analysis of various EU healthcare platforms, particularly those that specialize in elderly care. Insieme not only builds upon the foundation laid by these prior initiatives but also aims to set a new standard in the accessibility and usability of healthcare information, thereby providing a more integrated and cohesive experience for its users.

Additional medical knowledge therefore comes from the Insieme database that was created from a national electronic and mobile project for smart cities. Including the Insieme platform into the GPS is described elsewhere and is not part of this paper.

The HealthReview web application is developed in Python using the Flask framework. This application allows users to upload one or several PDF files containing multiple research papers, processes these files to extract the individual papers, and then employs ChatGPT to generate reviews for each paper.

The application is initialized and configured to use an upload folder named 'uploads'. This folder is created if it does not exist to ensure that uploaded files have a designated storage location. To maintain file security, the application restricts uploads to PDF files by specifying an allowed file extension set.

The user interface of the application consists of an HTML form rendered by the index route. This form allows users to upload their PDF files through a file input and submit button. Client-side JavaScript enhances user experience by displaying the selected file name and showing a loading indicator upon form submission.

Upon file upload via the /submit route, the PDF is securely saved in the upload folder. The file function ensures that only files with a .pdf extension are accepted. Once the file is stored, the papers_strArr function extracts the text of individual papers from the PDF. This function searches for specific markers like "ABSTRACT" and "Author index" to identify the boundaries of each paper. Text is extracted from each page and grouped into separate papers based on these markers. If no papers are detected, an error message is returned. Otherwise, the program proceeds to each extracted paper.

The API Call function facilitates the interaction with the ChatGPT model. It sends the text of each paper to the ChatGPT API, along with a predefined prompt, and retrieves the model's generated review. These reviews are stored in a global dictionary RESULTS with unique identifiers for each paper.

The */paper/<paper_id>* route enables users to view the generated reviews. When a specific paper ID is requested, the corresponding review content is fetched from the RESULTS dictionary and displayed using the paper.html template. Navigation links allow users to browse through the list of papers and view their respective reviews.

HTML templates, including form.html and paper.html, provide the user interface for uploading files, displaying instructions, and viewing individual paper reviews. The form.html template includes the file upload form and instructions for use, while the paper.html template displays individual paper reviews and includes navigation for switching between different papers.

This program integrates Flask for web functionality, PDFMiner for PDF text extraction, and OpenAI's ChatGPT-4o for generating paper reviews. This combination offers a streamlined and efficient method for reviewing conference papers, potentially enhancing the speed and consistency of the review process, which is crucial in managing the increasing volume of research submissions.

HealthReview accepts input in the form of any language and similarly outputs replies in any language supported by ChatGPT-40.

3 Results

The developed web application HealthReview was tested using a dataset of conference papers to evaluate its performance and efficiency in generating reviews. An example of the application's output is shown below.

Title: Anonymous in this paper
Authors: Anonymous
Abstract:
The article focuses on physical inactivity, which increases the risk of chronic diseases, especially cardiovascular diseases, among the adult population as well as among children and youth. Based on data from the Institute for Health Metrics and Evaluation, the article reports relative risks of cardiovascular disease by age group, gender, and country. In doing so, we focus on the comparison between Slovenia and the world average. Based on the analysis, we conclude that cardiovascular diseases impact people in different age segments; as such they become one of the important risk factors through aging. Also, physical activity importantly influences behavioral, nutrition, and other factors, which contribute to cardiovascular diseases. For that reason, it is very important to systematically analyze and monitor these factors to contribute to better guidelines and measures of policymakers.

Overall evaluation: 1 - Accept with minor changes
(All options: 2 - Accept as is | 1 - Accept with minor changes | 0 - Accept with major changes | -2 - Reject)
The paper provides valuable insights into the relationship between physical inactivity and cardiovascular diseases, particularly in different age segments. However, some minor improvements could elevate the presentation and clarity of the findings.
Overall quality: 7/10.
The scientific quality is solid, but minor methodological limitations and the need for more comprehensive data analysis slightly reduce its impact.
Presentation: 3 – Routine
(All options: 5 - Crystal-clear and particularly attractive | 4 - Clear and attractive | 3 - Routine | 2 - Not very clear and/or attractive | 1 - Unclear and/or particularly unattractive.)
The presentation is well-structured but lacks visual aids and a more engaging writing style.
Language modifications: Minor grammatical mistakes. The paper includes some minor grammatical mistakes that should be fixed before publication.
Proposed modifications: Correct "prerana" to "prehrana" in the "REZULTATI" section. Ensure consistency in the use of terms such as "risk factors" and "prevalence" across sections.
Improve the translation and readability of some sentences for better clarity.
Consider adding graphical representations of data to enhance interpretability.
Suggestion: Use tools like Grammarly or GPTs.
Reviewer's confidence: 4 – High
(All options: 5 - Expert | 4 - High | 3 - Medium | 2 - Low | 1 - None)
Excellence: 3 - Definite but modest methodological novelty and/or good application
(All options: 5 - Important methodological novelty | 4 - Definite methodological novelty | 3 - Definite but modest methodological novelty and/or good application | 2 - Debatable methodological novelty and/or acceptable application | 1 - No methodological novelty and poor application)
Impact: 3 - Used/useful and moderately mature
(All options: 5 - Highly impactful application and/or very widely used, at least moderately mature | 4 - Important application and/or widely used, at least moderately mature | 3 - Used/useful and moderately mature | 2 - Used/useful in limited cases and/or immature | 1 - No practical use)

The HealthReview web application successfully generated reviews for multiple conference papers. Table 1 represents a summary of the evaluations for the conference section.

4 Discussion

The results of our study demonstrate the promising potential of integrating AI, specifically ChatGPT, into the peer review process for healthcare and medical conference papers. The application we developed effectively generated insightful reviews, which were evaluated against traditional human-generated reviews for quality and consistency. While the overall performance of the AI-based review system was favorable, several considerations and implications warrant further discussion. First, the ability of ChatGPT to process and analyze large volumes of text rapidly offers a significant advantage in terms of efficiency. This is particularly beneficial in the context of increasing research output and the growing burden on peer reviewers. However, while the AI-generated reviews were generally accurate and aligned with human assessments, there were instances where the feedback provided by ChatGPT lacked depth, particularly in areas requiring domain-specific expertise. This highlights a limitation of current AI technologies, where the ingenuity, flexibility and expertise of human reviewers are still crucial.

It is not clear to what extent the automatic reviewing system applies to all domains, not only healthcare. In any case, when adopting a specific form of review, the instructions as part of the prompt programming should be modified. In practical terms, it should also be noticed that the API GPT call is not free of charge.

Additionally, the application of AI in the peer review process raises ethical concerns, particularly around the potential for bias, the risk of plagiarism, and the integrity of the review process. Although ChatGPT can streamline the review process, these

tools must be used as supplements rather than replacements for human reviewers. Maintaining a balance between AI efficiency and human oversight is critical to preserving the integrity and quality of academic peer reviews.

Moreover, the reliance on AI for academic tasks necessitates continuous monitoring and updates to the AI models to ensure accuracy, relevance, and fairness. Future developments should focus on enhancing the contextual understanding of AI tools like ChatGPT to better mimic the critical thinking and analytical capabilities of human reviewers.

Table 1: Summary of Reviews

Paper number	Overall evaluation	Overall quality	Presentation	Language (grammatical mistakes)
1	2/4	6/10	2/5	Quite some
2	3/4	7/10	3/5	Minor
3	3/4	7/10	3/5	Some
4	3/4	7/10	3/5	Quite some
5	3/4	7/10	3/5	Some
6	3/4	7/10	3/5	Minor
7	3/4	8/10	3/5	Minor
8	3/4	8/10	3/5	Some
9	3/4	7/10	3/5	Minor
10	3/4	7/10	3/5	Minor
11	3/4	8/10	4/5	Some
12	3/4	7/10	3/5	Minor
13	4/4	9/10	4/5	Minor

In conclusion, the inclusion of HealthReview, i.e. an additional automated review layer introduces several advantages, such as increased objectivity and the potential to generate supplementary suggestions, further enriching the review process. AI tools like ChatGPT offer substantial potential to enhance the peer review process. However, their successful integration requires careful implementation and continuous evaluation to effectively address inherent challenges and ensure that these tools make a meaningful contribution to academic research. In any case, the automatic review by HealthReview or any other review tool should be marked in a way explicitly denoting the source and type of the reviewing tool.

REFERENCES

[1] R. Garg, V. L. Urs, A. A. Agrawal, S. K. Chaudhary, V. Paliwal, and S. K. Kar. 2023. Exploring the role of chatgpt in patient care (diagnosis and treatment) and medical research: a systematic review. Health Promotion Perspectives, 13, 183–191. https://hpp.tbzmed.ac.ir/Article/HPP_19952_20230501034121

[2] P. Y. Lee, H. Salim, A. Abdullah, and C. H. Teo. 2023. Use of chatgpt in medical research and scientific writing. Malaysian Family Physician, 18, 58. https://www.e-mfp.org/ojs3/index.php/MFP/article/view/1213

[3] F. Tustumi, N. Andreollo, and J. E. Aguilar-Nascimento. 2023. Future of language models in healthcare: the role of chatgpt. Arquivos Brasileiros de Cirurgia Digestiva: ABCD, 36. https://doi.org/10.1590/0102-672020230002e1800

[4] Ö. Aydın and E. Karaarslan. 2022. Openai chatgpt generated literature review: digital twin in healthcare. In Emerging Computer Technologies 2. Ö. Aydın, editor. İzmir Akademi Dernegi, 22–31. https://doi.org/10.5281/zenodo.7094305S.

[5] S. Ruksakulpiwat, A. Kumar, and A. Ajibade. 2023. Using chatgpt in medical research: current status and future directions. Journal of Multidisciplinary Healthcare, 16, 1513–1520 https://doi.org/10.2147/JMDH.S419830.

[6] M. Sallam. 2023. Chatgpt utility in healthcare education, research, and prac- tice: systematic review on the promising perspectives and valid concerns. Healthcare, 11. doi: 10.3390/healthcare11060887 D. Solomon, K. Allen, P. Katz, A. Sawalha, and E. Yelin. 2023. Chatgpt, et al... artificial intelligence, authorship, and medical publishing. Arthritis & Rheumatology, 75. doi: 10.1002/art.42497

[7] O. Temsah et al. 2023. Overview of early chatgpt's presence in medical literature: insights from a hybrid literature review by chatgpt and human experts. Cureus, 15. doi: 10.7759/cureus.37281

[8] S. Wang, H. Scells, B. Koopman, and G. Zuccon. 2023. Can chatgpt write a good boolean query for systematic review literature search? In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. doi: 10.1145/3539618.3591703

[9] Biswas, D. Dobaria, and H. L. Cohen. 2023. Chatgpt and the future of journal reviews: a feasibility study. The Yale Journal of Biology and Medicine, 96, 415–420. https://pubmed.ncbi.nlm.nih.gov/37407690/

[10] D. Solomon, K. Allen, P. Katz, A. Sawalha, and E. Yelin. 2023. Chatgpt, et al... artificial intelligence, authorship, and medical publishing. Arthritis & Rheumatology, 75. doi: 10.1002/art.42497

[11] J. Verharen. 2023. Chatgpt identifies gender disparities in scientific peer review. eLife, 12. https://doi.org/10.7554/eLife.88173