SHORT COMMUNICATION

Enhancing quotation accuracy assessment with Chatpdf – a game-changer for a century-old conundrum

Muhammad Talal Ibrahim[1], Cole Vincent Veliky[2], Syeda Sadia Fatima[3], Zahra Hoodbhoy[4], Shahryar Noordin[5]

Abstract

Quotation errors compromise the reliability of published medical literature, but assessing these errors is very time-intensive. The current observational cross-sectional study was planned to assess the accuracy and speed of detecting quotation errors using the ChatPDF programme that is powered by artificial intelligence. Of the 398 quotations assessed, 310(77.9%) were fully supported, while 88(22.1%) had errors. Among the former quotations, the ChatPDF programme could comprehensively highlight 210(67.7%) PDFs, provided comprehensive answers in 262(84.5%), and was completely helpful in 248(80%) cases. In contrast, for erroneous quotations, the corresponding values were 36(40.9%), 54(61.4%) and 43(48.9%), respectively ($p<0.001$). ChatPDF was found to have immense potential to revolutionise quotation error assessments.

Keyword: Artificial intelligence, Natural language processing, ChatPDF, Referencing error, Quotation error.

DOI: https://doi.org/10.47391/JPMA.AKU-10Surg-28

Introduction

Researchers substantiate and contextualise their work by citing the existing literature. The 'publish-or-perish' pressure leads to errors in the published literature, including referencing errors that have been a concern since 1916 and continue to be a significant problem[1,2]. These compromise the reliability of the published medical literature used to secure grant funding, provide rationale for clinical trials, and make clinical practice guidelines.

Referencing errors are categorised into citation and quotation errors. Automated reference management softwares have largely addressed citation errors[1]. However, quotation errors are still prevalent and challenging to reduce primarily because it is a time-intensive task to assess quotation accuracy[3]. Most solutions proposed to reduce quotation errors[1,4] have not been effective. A major reason is the time-intensive nature of most solutions. Therefore, more time-efficient and practical solutions are needed for effective results.

ChatPDF, powered by ChatGPT-3.5, which is a programme based on artificial intelligence (AI), allows interaction with a portable document format (PDF) using a natural language chatbox. It appears to be a solution to the century-old problem of quotation errors. To our knowledge, there is a lack of data on the evaluation of ChatPDF related to quotation accuracy. The current study was planned to fill the gap in literature by assessing the accuracy and time-efficiency of ChatPDF in detecting quotation errors.

Materials and Methods

The observational, cross-sectional study was conducted at the Aga Khan University Medical College from January 2024 to July 2024. After exemption from the institutional ethics review committee as no identifiable patient data was accessed, the study sample comprised the top-four cited articles published between 2013 and 2023 in the Journal of Bone and Joint Surgery and the Journal of the American Medical Association – Surgery (JAMA Surgery). The articles were identified using the Scopus database. The term 'quotation' was defined as a statement in the text with a citation. Quotations citing websites or software were excluded because ChatPDF only interacts with PDFs.

PDF files were uploaded for each referenced article to ChatPDF, and 'quotation' was used as the prompt. Although a predetermined standardised format for the prompts was created, it could not be used because of a wide variation in sentence structure and placement of in-text citations (mid-sentence versus end-of-sentence). Moreover, prompts were tweaked to make them a suitable natural language input for the ChatPDF.

ChatPDF provides an answer in the chatbox, and, for

[1,5]Section of Orthopedics, Department of Surgery, Aga Khan University Hospital, [2]Ohio State University College of Medicine, Columbus, Ohio, USA [3]Department of Biological and Biomedical Sciences, Aga Khan University, [4]Department of Pediatrics and Child Health, Aga Khan University, Karachi, Pakistan

Correspondence: Muhammad Talal Ibrahim.
Email: mtalalibrahim@gmail.com
ORCID ID: 0000-0002-6687-0418

some answers, it provides highlights relevant to the text in the PDF. To assess the quotation accuracy, both the answers and the highlights were examined. If these were insufficient, the reviewer read the entire PDF. The key for assessing quotation accuracy was adapted from previous studies [3,5-7]. In line with certain studies [3,6], quotations with string citations were considered fully supported if each cited reference supported at least a part of the quotation.

The time taken to mark quotations and download PDFs was noted, and a stopwatch was used to record the time taken to assess quotation accuracy using ChatPDF. This included the time spent manually verifying the accuracy of ChatPDF's responses. To estimate the time taken for the manual assessment of quotation accuracy without ChatPDF, authors of previous studies[3,7,8] on the topic were contacted. Data was analysed using Stata 17. Descriptive statistics were calculated and hypothesis testing was done using the chi-square test. P<0.05 was considered significant.

Results

Of the 430 quotations identified from the 8 published studies, 32(7.4%) were excluded, and 398(92.6%) were assessed. Of them, 310(77.9%) were fully supported, while 88(22.1%) had errors (Figure). Among the former quotations, the ChatPDF programme could comprehensively highlight 210(67.7%) PDFs, provided comprehensive answers in 262(84.5%), and was completely helpful in 248(80%) cases. In contrast, for erroneous quotations, the corresponding values were 36(40.9%), 54(61.4%) and 43(48.9%), respectively (p<0.001) (Table).

Table: Citation style and ChatPDF's performance based on accuracy of quotations.

	Fully Supported Quotation 310 (77.9%)	Erroneous Quotation 88 (22.1%)	p Value
Citation Style			
Single Citation	122 (39.4%)	28 (31.8%)	0.198
String Citation	188 (60.7%)	60 (68.2%)	
ChatPDF - Highlighting			
Comprehensive	210 (67.7%)	36 (40.9%)	< 0.001
Inadequate	87 (28.1%)	26 (29.6%)	
Not applicable	13 (4.2%)	26 (29.6%)	
ChatPDF – Chat Answer			
Comprehensive	262 (84.5%)	54 (61.4%)	< 0.001
Incomplete	27 (8.7%)	20 (22.7%)	
Incorrect	21 (6.8%)	14 (15.9%)	
ChatPDF – Guidance to Correct Assessment			
Completely Helpful	248 (80.0%)	43 (48.9%)	< 0.001
Partially Helpful	38 (12.3%)	28 (31.8%)	
Not Helpful	38 (12.3%)	28 (31.8%)	

Using ChatPDF, quotation accuracy was assessed at a mean (SD) rate of 5.3 (1.3) minutes per quotation per

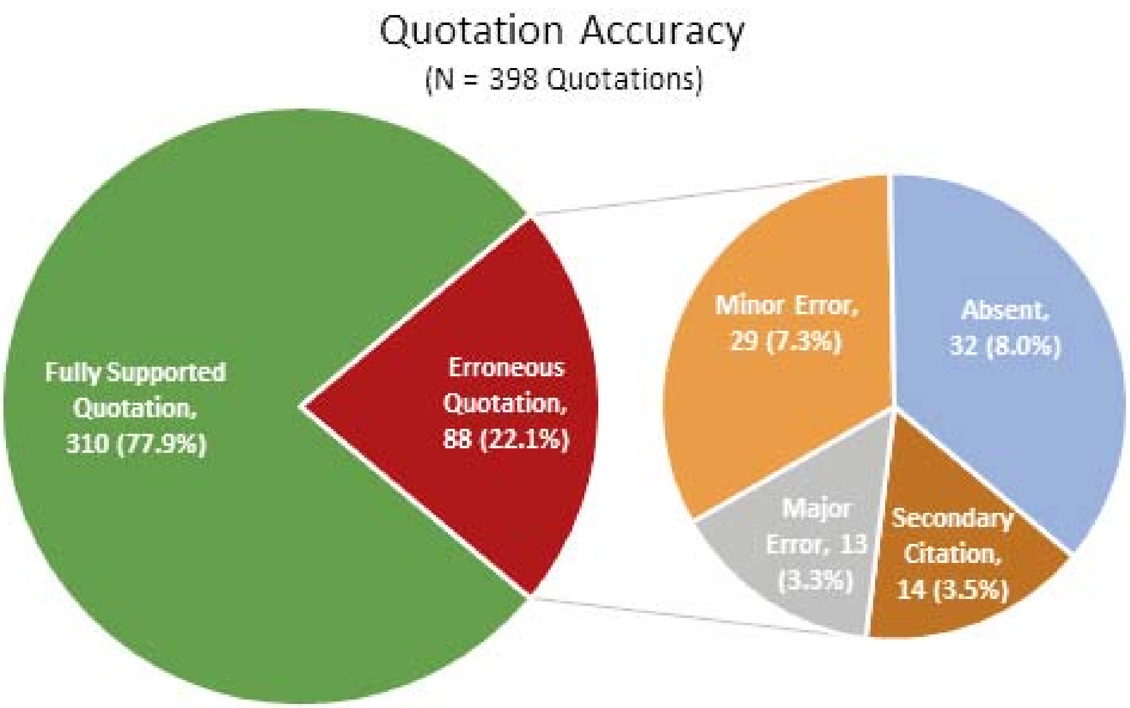

Figure: Quotation accuracy assessment using ChatPDF.

reviewer (m/q/r). Based on the rough estimates provided by authors of the previous studies, it took an average (SD) of 25.1 (5.5) m/q/r to manually assess the quotation accuracy without ChatPDF. Hence, ChatPDF led to an estimated 19.8(79%)m/q/r reduction in quotation accuracy assessment time.

Discussion

To our knowledge, the current study is the first to evaluate the use of the AI-based ChatPDF tool for quotation accuracy assessment. The quotation error rates in the study are consistent with previous findings[2,9]. This suggests little has changed since Lacey et al. first estimated quotation error rates in 1985[5]. Despite multiple publications repeatedly raising the same issue in the interim period[9], this lack of improvement is concerning. Although previous solutions[1,4] have failed to address the elements of time and effort required to assess quotation accuracy, ChatPDF saved an estimated 79% of the previously required time.

ChatPDF is an effective tool for assessing quotation accuracy, particularly helpful for lengthy citations. The ChatPDF's answers in the chatbox seemed more accurate than its highlighting function. However, users should exercise caution in relying solely on ChatPDF-generated answers, since there is no data on the accuracy or reliability of these answers yet. The current study observed cases in which the responses contradicted themselves or quoted non-existent phrases. Hence, until a more reliable version is available, verifying ChatPDF's answers is advised to avoid these 'hallucinations.' This observation is in congruence with other studies reporting errors in ChatGPT's answers[10], the AI tool used in ChatPDF.

ChatPDF also seemed to struggle when numbers were derived from data in the PDF, but were not directly copy-pasted. Moreover, ChatPDF also struggled to read the data from tables and graphs. However, it was able to read captions and direct the researchers to the table or figure if the caption was sufficiently descriptive. It also struggled with vague quotations, like "our results were similar to the literature", without mentioning any specific results.

ChatPDF has the potential to revolutionise academic publishing with an estimated 79% reduction in time required to check quotation accuracy. In the short term, it can help in two ways. First, it can accelerate the research done on quotation errors in literature by reducing the time required, helping researchers and editors expand their understanding of quotation errors. Second, journals can start pilot programmes to start quotation accuracy checks during the manuscript review process. Editorial staff can mark potentially erroneous quotations, which can be sent for verification to authors. Moreover, peer-reviewers can be asked to highlight the 'critical' quotations which form the basic rationale or explanation of the manuscript, and these critical quotations can be fact-checked. This approach should not take >2 minutes of reviewers' time, 53 minutes of editorial staff time, and 14 minutes of authors' time, based on the 25% average error rate in current literature [9]. In the long term, one can hope to have more automated quotation-checking tools, similar to automated plagiarism-checking softwares.

There current study has a few limitations. First, it only checked accuracy of the referenced article, not the article cited by that article. Second, a single reviewer checked the references. However, since the current study was primarily a proof-of-concept study, these limitations were beyond the scope of the primary aim. Future studies would benefit from using updated versions of the AI tool, including a larger sample size and blinded double scoring for each citation. Future AI tools should incorporate image recognition, the ability to read tables, and mathematical capabilities in analysing and answering user prompts.

Conclusion

Tools such as ChatPDF could have immense potential to revolutionise quotation error assessments, reducing assessment time by an estimated 79% compared to the traditional methods.

Disclaimer: None.

Conflict of Interest: None.

Source of Funding: None.

References

1. Agarwal A, Arafa M, Avidor-Reiss T, Hamoda TAA, Shah R. Citation Errors in Scientific Research and Publications: Causes, Consequences, and Remedies. World J Mens Health 2023;41:461-5. doi: 10.5534/wjmh.230001.
2. Mogull SA. Accuracy of cited "facts" in medical research articles: A review of study methodology and recalculation of quotation error rate. PLoS One 2017;12:e0184727. doi: 10.1371/journal.pone.0184727.
3. Smith N, Cumberledge A. Quotation errors in general science journals. Proc. A 2020;476: 20200538 .doi: 10.1098/rspa.2020.0538 2020;476.
4. Lin TY, Hung TM. How to Reduce Errors and Improve Transparency by Using More Precise Citations. Front Cardiovasc Med 2022;9:e866279. doi: 10.3389/fcvm.2022.866279..
5. de Lacey G, Record C, Wade J. How accurate are quotations and references in medical journals? Br Med J (Clin Res Ed) 1985;291:884-6. doi: 10.1136/bmj.291.6499.884.
6. Gazendam A, Cohen D, Morgan S, Ekhtiari S, Ghert M. Quotation Errors in High-Impact-Factor Orthopaedic and Sports Medicine Journals. JB JS Open Access 2021;6:e21.00019. doi:

10.2106/JBJS.OA.21.00019.
7. Davids JR, Weigl DM, Edmonds JP, Blackhurst DW. Reference accuracy in peer-reviewed pediatric orthopaedic literature. J Bone Joint Surg Am 2010;92:1155-61. doi: 10.2106/JBJS.I.00063.
8. Choi J, Gupta A, Kaghazchi A, Htwe TS, Baiocchi M, Spain DA. Citation Inaccuracies in Influential Surgical Journals. JAMA Surg 2021;156:791-2. doi: 10.1001/jamasurg.2021.1445.
9. Jergas H, Baethge C. Quotation accuracy in medical journal articles-a systematic review and meta-analysis. PeerJ 2015;3:e1364. doi: 10.7717/peerj.1364.
10. Metze K, Morandin-Reis RC, Lorand-Metze I, Florindo JB. The Amount of Errors in ChatGPT's Responses is Indirectly Correlated with the Number of Publications Related to the Topic Under Investigation. Ann Biomed Eng 2023;51:1360-1. doi: 10.1007/s10439-023-03205-1.