
Short Paper

Use of ChatGPT to Explore Gender and Geographic Disparities in Scientific Peer Review

Paul Sebo, MSc, MD

University of Geneva, Geneva, Switzerland

Corresponding Author:

Paul Sebo, MSc, MD

University of Geneva

Rue Michel-Servet 1

Geneva, 1211

Switzerland

Phone: 41 223794390

Email: paul.seboe@unige.ch

Abstract

Background: In the realm of scientific research, peer review serves as a cornerstone for ensuring the quality and integrity of scholarly papers. Recent trends in promoting transparency and accountability has led some journals to publish peer-review reports alongside papers.

Objective: ChatGPT-4 (OpenAI) was used to quantitatively assess sentiment and politeness in peer-review reports from high-impact medical journals. The objective was to explore gender and geographical disparities to enhance inclusivity within the peer-review process.

Methods: All 9 general medical journals with an impact factor >2 that publish peer-review reports were identified. A total of 12 research papers per journal were randomly selected, all published in 2023. The names of the first and last authors along with the first author's country of affiliation were collected, and the gender of both the first and last authors was determined. For each review, ChatGPT-4 was asked to evaluate the "sentiment score," ranging from -100 (negative) to 0 (neutral) to +100 (positive), and the "politeness score," ranging from -100 (rude) to 0 (neutral) to +100 (polite). The measurements were repeated 5 times and the minimum and maximum values were removed. The mean sentiment and politeness scores for each review were computed and then summarized using the median and interquartile range. Statistical analyses included Wilcoxon rank-sum tests, Kruskal-Wallis rank tests, and negative binomial regressions.

Results: Analysis of 291 peer-review reports corresponding to 108 papers unveiled notable regional disparities. Papers from the Middle East, Latin America, or Africa exhibited lower sentiment and politeness scores compared to those from North America, Europe, or Pacific and Asia (sentiment scores: 27 vs 60 and 62 respectively; politeness scores: 43.5 vs 67 and 65 respectively, adjusted $P=.02$). No significant differences based on authors' gender were observed (all $P>.05$).

Conclusions: Notable regional disparities were found, with papers from the Middle East, Latin America, and Africa demonstrating significantly lower scores, while no discernible differences were observed based on authors' gender. The absence of gender-based differences suggests that gender biases may not manifest as prominently as other forms of bias within the context of peer review. The study underscores the need for targeted interventions to address regional disparities in peer review and advocates for ongoing efforts to promote equity and inclusivity in scholarly communication.

(*J Med Internet Res* 2024;26:e57667) doi: [10.2196/57667](https://doi.org/10.2196/57667)

KEYWORDS

Africa; artificial intelligence; discrimination; peer review; sentiment analysis; ChatGPT; disparity; gender; geographic; global south; inequality; woman; assessment; researcher; communication; consultation; gender bias

Introduction

The peer-review process plays a pivotal role in validating the quality and integrity of scholarly papers. With an increasing

emphasis on transparency and accountability, some journals adopted the practice of publishing peer-review reports alongside papers.

Sentiment analysis, the identification and categorization of opinions, attitudes, and emotions conveyed in text data as positive, negative, or neutral [1], finds ChatGPT (OpenAI) standing out with unique advantages compared to “traditional methods” [1-5]. Unlike lexicon-based approaches that may overlook nuanced expressions, ChatGPT understands natural language, capturing subtle nuances and context-specific sentiments. Unlike machine learning methods requiring labeled datasets, ChatGPT adapts to diverse domains without explicit training, reducing costs and time. Its humanlike responses facilitate intuitive sentiment interpretation. Additionally, artificial intelligence (AI)-driven sentiment analysis, including ChatGPT, ensures efficient, scalable (able to accommodate increasing data volumes without significant performance degradation), consistent, objective (devoid of human biases and preconceptions), and adaptable analyses. ChatGPT showed promise for sentiment analysis in several recent studies [6-10].

Verharen [11] showed that ChatGPT was accurate in determining sentiment and politeness scores in peer reviews of scientific papers. The study also showed gender inequalities, with woman authors receiving less polite reviews than men. The study, limited to papers from a single journal (*Nature Communications*), aligns with the broader issue of gender discrimination in academic medicine [12-17]. While sentiment and politeness metrics may not directly measure bias, they serve as useful proxies for identifying potential biases in peer review.

Biased reviewers may exhibit tendencies toward overly positive or negative sentiments, as well as varying levels of politeness toward authors based on factors such as gender or geographic region.

Building upon Verharen’s [11] work, ChatGPT-4 was used to quantitatively assess sentiment and politeness in peer-review reports from 9 high-impact medical journals, exploring gender and geographical disparities to enhance inclusivity within the peer review process. By leveraging the capabilities of AI, this study sheds light on the potential of AI technologies to mitigate biases and promote fairness in scholarly communication.

Methods

Selection of Journals, Papers, and Peer-Review Reports

The Clarivate and ASAPbio websites were searched to identify all 9 general medical journals with a journal citation reports impact factor >2 that publish peer-review reports (Table 1). None of these journals uses double-blind peer review. Using simple randomization, which involves randomly selecting papers using a random number generator, 12 research papers per journal were selected, all published in 2023, and all peer-review reports from the initial round for these 108 papers were retrieved. This sample size was chosen to ensure robust and manageable analysis.

Table 1. List of high-impact general medical journals included in this cross-sectional study using ChatGPT-4 to quantitatively assess sentiment and politeness in 291 peer-review reports for 108 papers published in 2023, as well as number of papers and reviews per journal, and median sentiment and politeness scores per journal.

Journal	2022 impact factor, n	Papers (n=108), n (%)	Reviews (n=291), n (%)	Sentiment score, median (IQR)	Politeness score, median (IQR)
<i>BMJ</i>	107.7	12 (11.1)	51 (17.5)	68 (53-78)	73 (60-80)
<i>PLOS Medicine</i>	15.8	12 (11.1)	41 (14.1)	60 (50-73)	70 (62-78)
<i>BMC Medicine</i>	9.3	12 (11.1)	29 (10.0)	63 (48-70)	73 (60-78)
<i>Journal of Clinical Medicine</i>	3.9	12 (11.1)	31 (10.7)	43 (17-70)	50 (25-67)
<i>Diagnostics</i>	3.6	12 (11.1)	28 (9.6)	37 (3-60)	57 (28.5-68.5)
<i>Journal of Personalized Medicine</i>	3.4	12 (11.1)	27 (9.3)	57 (17-70)	68 (27-82)
<i>BMJ Open</i>	2.9	12 (11.1)	27 (9.3)	63 (40-72)	60 (45-70)
<i>BMC Primary Care</i>	2.9	12 (11.1)	25 (8.6)	57 (10-67)	55 (30-70)
<i>Medicina</i>	2.6	12 (11.1)	32 (11.0)	48.5 (20-65)	57 (43.5-66)

Data Collection and Gender Determination

The names of the first and last authors along with the first author’s country of affiliation were collected, and the gender of both the first and last authors was determined. The authors’ genders were categorized in 2 steps. The gender was determined based on names alone, classifying names as man or woman accordingly. For authors whose gender could not be inferred from their names, professional networks and university websites were searched for photos or text containing gender-specific pronouns. This method enabled to assign genders to all authors. A gender detection tool (Gender API; Markus Perl IT Solutions) was also used to confirm the classifications [18]. Gender API demonstrated high accuracy in previous studies [19]. Both

approaches yielded similar results, with high agreement between them (first authors: percentage agreement=0.9725, Cohen κ =0.9450; last authors: percentage agreement=0.9691, Cohen κ =0.9295).

Sentiment and Politeness Scores

For each review, ChatGPT-4 was asked to evaluate the “sentiment score,” ranging from –100 (negative) to +100 (positive), and the “politeness score,” ranging from –100 (rude) to +100 (polite). The measurements were repeated 5 times and the minimum and maximum values were removed. The sentiment score measures how favorable the review is, and the politeness score how polite a review’s language is. The same prompt as Verharen [11] was used:

Below you will find a scientific peer review. Can you score this peer review on the sentiment, on a scale from -100 (negative) to 0 (neutral) to 100 (positive), and politeness of language use, on a scale of -100 (rude) to 0 (neutral) to 100 (polite)?

All the data were collected in January 2024. The data associated with this paper are available in the Open Science Framework [20]. The database is provided as a “dta” file for use with Stata. [Multimedia Appendix 1](#) provides a detailed list and description of the variables included in the Stata file.

Statistical Analyses

The mean sentiment and politeness scores for each review were computed and rounded to the nearest whole number. These scores were then summarized using the median and IQR, both overall and stratified by journal, gender, and affiliation. Affiliation countries were categorized into 3 regions (North America, Europe, or Pacific; Asia; and Latin America, Middle East, or Africa), following prior research [21]. Comparisons were conducted using Wilcoxon rank-sum tests (for gender) and Kruskal-Wallis rank tests (for affiliation). Negative binomial regressions were performed, adjusting for journal, affiliation, and intracluster correlation within papers [22,23]. To accommodate the requirement of nonnegative outcome variables in negative binomial regressions, 100 were added to the sentiment and politeness scores, resulting in a scale from 0 to 200. Quadratic weighted agreement coefficients were calculated to assess the agreement between the 3 sentiment and politeness score measurements. Fleiss κ was used instead of Cohen κ due to the presence of more than 2 measurements [24]. All analyses were conducted using Stata (version 15.1; StataCorp).

Ethical Considerations

Since this study did not involve the collection of personal health-related data, it did not require ethical review, according to the current Swiss law.

Results

There were 291 reviews for the 108 papers selected for the study ([Multimedia Appendix 2](#)). Men were the first and last authors of 61 (56.5%) and 75 (69.4%) papers, respectively. The 5 most represented countries of affiliation were the United Kingdom (n=14), Germany (n=13), the United States (n=10), China (n=9), and Italy (n=6). The 3 main regions of affiliation were Western Europe (n=56), Asia (n=16), and North America (n=15).

Overall, the median sentiment and politeness scores were 58 (IQR 30-72; range -70 to 90) and 63 (IQR 47-75; range -73 to 92), respectively, but there were notable variations by the journal ([Table 1](#)). The 3 journals with the highest impact factor tended to have higher sentiment and politeness scores. There was no significant difference in scores between men and women (all $P>.05$; [Table 2](#)). The results were almost identical when using Gender API to determine gender (first authors: median sentiment or politeness scores=58, IQR 37-72 and 65, IQR 47-77 for women, 57.5, IQR 27-70 and 63, IQR 47-73 for men; last authors: median sentiment or politeness scores=55, IQR 33-68 and 63, IQR 50-75 for women, 58, IQR 27-72 and 63, IQR 45-75 for men). By contrast, papers authored by scholars from countries in the Middle East, Latin America, or Africa exhibited significantly lower sentiment and politeness scores compared to those from the other 2 regions, with differences exceeding 30 and 20 points in absolute value, respectively (adjusted $P=.02$; [Table 2](#)).

In light of the *BMJ*'s notable impact factor compared to the other journals in the study, additional analyses were conducted by excluding the 51 peer-review reports for *BMJ*. The results remained consistent ([Multimedia Appendix 3](#)). The interrater agreement between the 3 measurements was high (sentiment scores: percentage agreement=0.9958, 95% CI 0.9954-0.9962; Fleiss κ =0.9496, 95% CI 0.9395-0.9598; politeness scores: percentage agreement=0.9962, 95% CI 0.9958-0.9966; Fleiss κ =0.9463, 95% CI 0.9316-0.9610; all $P<.001$).

Table 2. Associations between sentiment or politeness scores and first or last authors' gender and first authors' affiliation in this cross-sectional study using ChatGPT-4 to quantitatively assess sentiment and politeness in 291 peer-review reports for 108 papers published in 2023 in 9 high-impact general medical journals.

Variable	Papers (n=108), n (%)	Reviews (n=291), n (%)	Sentiment score, median (IQR)	Crude <i>P</i> value ^a	Adjusted <i>P</i> value ^b	Politeness score, median (IQR)	Crude <i>P</i> value ^a	Adjusted <i>P</i> value ^b
First authors' gender				.49	.48		.37	.68
Woman	47 (43.5)	127 (43.6)	58 (33-72)			65 (47-77)		
Man	61 (56.5)	164 (56.4)	57.5 (27-70)			63 (48.5-73)		
Last authors' gender				.52	.88		.74	.86
Woman	33 (30.6)	91 (31.3)	57 (33-68)			63 (50-75)		
Man	75 (69.4)	200 (68.7)	60 (27.5-72)			63 (45-75)		
First authors' affiliation				.001	.02 ^c		<.001	.02 ^d
North America, Europe, and Pacific	82 (75.9)	220 (75.6)	60 (33-72)			67 (53-77)		
Asia	16 (14.8)	43 (14.8)	62 (40-70)			65 (50-73)		
Middle East, Latin America, and Africa	10 (9.3)	28 (9.6)	27 (-3 to 55)			43.5 (17.5-57)		

^aWilcoxon rank-sum test (for gender) and Kruskal-Wallis equality-of-populations rank test (for affiliation).

^bMultivariable negative binomial regression, adjusted for journal, affiliation, and intracluster correlation within papers (for first or last authors' gender), and adjusted for journal and intracluster correlation within papers (for first authors' affiliation).

^cIncidence rate ratio: Asia versus Middle East, Latin America, and Africa: 1.27 (95% CI 1.06-1.51); North America, Europe, and Pacific versus Middle East, Latin America, and Africa: 1.23 (95% CI 1.02-1.47).

^dIncidence rate ratio: Asia versus Middle East, Latin America, and Africa: 1.30 (95% CI 1.07-1.57); North America, Europe, and Pacific versus Middle East, Latin America, and Africa: 1.27 (95% CI 1.04-1.54).

Discussion

Principal Findings

ChatGPT-4 was used to analyze sentiment and politeness in 291 peer-review reports from 9 general medical journals. The study unveiled notable regional disparities, with papers from the Middle East, Latin America, or Africa demonstrating significantly lower scores, while no discernible differences were observed based on the authors' gender.

Comparison With Existing Literature

The gender disparities experienced by women in academic medicine are widely recognized [12-17]. Consequently, significant gender discrepancies were anticipated in this study to the detriment of women. The absence of discernible differences between genders contrasts with Verharen's [11] findings, where woman first authors typically received less polite reviews compared to men. Importantly, the study focused on a different discipline (neuroscience) and was confined to a single journal (*Nature Communications*).

The regional disparities highlighted in the study resonate with prior research illustrating the challenges faced by researchers, particularly those from countries in the Global South [21,25]. These findings align with those of earlier studies [26-28] and must be understood within broader sociocultural, economic, and institutional contexts. Factors such as limited funding opportunities, language barriers, and cultural differences may contribute to biases against authors from underrepresented regions. Additionally, institutional biases within academic

publishing systems may further exacerbate these discrepancies, highlighting the need for interventions to foster equity and inclusivity in scholarly discourse.

The observed disparities across geographic regions, contrasted with the absence of significant differences based on gender, could indicate that within the context of peer review, gender biases may not manifest as prominently or uniformly as other forms of bias, such as those influenced by geographic factors. In addition, existing diversity or inclusion initiatives within the scholarly community may have been more effective in mitigating gender disparities compared to other forms of bias.

Strengths and Limitations

This study built on the findings of Verharen [11], who demonstrated through several validation methods the accuracy of ChatGPT in estimating sentiment and politeness in peer-review reports, surpassing that of human evaluation and traditional lexicon-based language models. This study has several limitations, including its exclusive focus on high-impact general medical journals, reliance on binary gender determination without consideration for nonbinary or transgender identities, and uncertainty about gender or geographic distributions of rejected papers. Future research should explore alternative approaches for gender determination such as self-identification for accurately capturing gender diversity. In addition, the sample size (108 papers and 291 peer-review reports) is smaller than that of prior research by Verharen [11], potentially limiting the generalizability of the findings. The author also acknowledges the challenges

associated with using sentiment or politeness metrics to capture the nuanced biases inherent in peer review, and the dependence on algorithms like ChatGPT may introduce potential inaccuracies. Furthermore, manual scoring was not conducted for comparison. Finally, the observed association between sentiment and politeness metrics and affiliation regions suggests an alternative explanation: the possibility of higher scientific merit in papers from certain authors. The methodology lacks the capability to discern between the 2 hypotheses.

Conclusions

ChatGPT-4 demonstrated effectiveness in this study by consistently evaluating sentiment and politeness in peer-review reports. The study underscores the need for targeted interventions to address regional disparities in peer review and advocates for ongoing efforts to promote equity and inclusivity in scholarly communication.

Data Availability

The datasets generated and analyzed during this study are available in the Open Science Framework (OSF) repository [20].

Conflicts of Interest

None declared.

Multimedia Appendix 1

List and description of variables available in the Stata file uploaded to the Open Science Framework (OSF).

[\[DOCX File , 25 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

List of papers and reviews.

[\[XLSX File \(Microsoft Excel File\), 25 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

Associations between sentiment or politeness scores and first or last authors' gender and first authors' affiliation in this cross-sectional study using ChatGPT-4, to quantitatively assess sentiment and politeness in 240 peer review reports for 96 articles published in 2023 (same data as in Table 2 but without taking into account the 51 peer-review reports for the 12 articles published in BMJ).

[\[DOCX File , 16 KB-Multimedia Appendix 3\]](#)

References

1. Mehta P, Pandya DS. A review on sentiment analysis methodologies, practices and applications. *Int J Sci Technol Res.* 2020;9:601-609. [\[FREE Full text\]](#)
2. Bordoloi M, Biswas SK. Sentiment analysis: a survey on design framework, applications and future scopes. *Artif Intell Rev.* 2023;56:12505-12560. [\[FREE Full text\]](#) [doi: [10.1007/s10462-023-10442-2](https://doi.org/10.1007/s10462-023-10442-2)] [Medline: [37362892](https://pubmed.ncbi.nlm.nih.gov/37362892/)]
3. Cui J, Wang Z, Ho SB, Cambria E. Survey on sentiment analysis: evolution of research methods and topics. *Artif Intell Rev.* 2023;56:8469-8510. [\[FREE Full text\]](#) [doi: [10.1007/s10462-022-10386-z](https://doi.org/10.1007/s10462-022-10386-z)] [Medline: [36628328](https://pubmed.ncbi.nlm.nih.gov/36628328/)]
4. Wankhade M, Rao ACS, Kulkarni C. A survey on sentiment analysis methods, applications, and challenges. *Artif Intell Rev.* 2022;55(7):1-50. [doi: [10.1007/s10462-022-10144-1](https://doi.org/10.1007/s10462-022-10144-1)]
5. Islam T, Sheakh MA, Sadik MR, Tahosin MS, Foysal MMR, Ferdush J, et al. Lexicon and deep learning-based approaches in sentiment analysis on short texts. *JCC.* 2024;12(01):11-34. [doi: [10.4236/jcc.2024.121002](https://doi.org/10.4236/jcc.2024.121002)]
6. Fu Z, Hsu YC, Chan CS, Lau CM, Liu J, Yip PSF. Efficacy of ChatGPT in Cantonese sentiment analysis: comparative study. *J Med Internet Res.* 2024;26:e51069. [\[FREE Full text\]](#) [doi: [10.2196/51069](https://doi.org/10.2196/51069)] [Medline: [38289662](https://pubmed.ncbi.nlm.nih.gov/38289662/)]
7. Lossio-Ventura JA, Weger R, Lee AY, Guinee EP, Chung J, Atlas L, et al. A comparison of ChatGPT and fine-tuned open pre-trained transformers (OPT) against widely used sentiment analysis tools: sentiment analysis of COVID-19 survey data. *JMIR Ment Health.* 2024;11:e50150. [\[FREE Full text\]](#) [doi: [10.2196/50150](https://doi.org/10.2196/50150)] [Medline: [38271138](https://pubmed.ncbi.nlm.nih.gov/38271138/)]
8. Tabone W, de Winter J. Using ChatGPT for human-computer interaction research: a primer. *R Soc Open Sci.* 2023;10(9):231053. [\[FREE Full text\]](#) [doi: [10.1098/rsos.231053](https://doi.org/10.1098/rsos.231053)] [Medline: [37711151](https://pubmed.ncbi.nlm.nih.gov/37711151/)]
9. Kim S, Kim K, Wonjeong Jo C. Accuracy of a large language model in distinguishing anti- and pro-vaccination messages on social media: the case of human papillomavirus vaccination. *Prev Med Rep.* 2024;42:102723. [\[FREE Full text\]](#) [doi: [10.1016/j.pmedr.2024.102723](https://doi.org/10.1016/j.pmedr.2024.102723)] [Medline: [38659997](https://pubmed.ncbi.nlm.nih.gov/38659997/)]
10. Wang Z, Xie Q, Feng Y, Ding Z, Yang Z, Xia R. Is ChatGPT a good sentiment analyzer? A preliminary study. *arXiv.* Preprint posted online on April 10, 2023. [doi: [10.48550/arXiv.2304.04339](https://doi.org/10.48550/arXiv.2304.04339)]
11. Verharen JPH. ChatGPT identifies gender disparities in scientific peer review. *Elife.* 2023;12(RP90230). [\[FREE Full text\]](#) [doi: [10.7554/eLife.90230](https://doi.org/10.7554/eLife.90230)] [Medline: [37922198](https://pubmed.ncbi.nlm.nih.gov/37922198/)]

12. Richter KP, Clark L, Wick JA, Cruvinel E, Durham D, Shaw P, et al. Women physicians and promotion in academic medicine. *N Engl J Med*. 2020;383(22):2148-2157. [doi: [10.1056/NEJMsa1916935](https://doi.org/10.1056/NEJMsa1916935)] [Medline: [33252871](https://pubmed.ncbi.nlm.nih.gov/33252871/)]
13. Sebo P, de Lucia S, Vernaz N. Gender gap in medical research: a bibliometric study in Swiss university hospitals. *Scientometrics*. 2020;126(4):741-755. [doi: [10.1007/s11192-020-03741-w](https://doi.org/10.1007/s11192-020-03741-w)]
14. Hart KL, Perlis RH. Trends in proportion of women as authors of medical journal articles, 2008-2018. *JAMA Intern Med*. 2019;179(9):1285-1287. [FREE Full text] [doi: [10.1001/jamainternmed.2019.0907](https://doi.org/10.1001/jamainternmed.2019.0907)] [Medline: [31135815](https://pubmed.ncbi.nlm.nih.gov/31135815/)]
15. Sebo P, Clair C. Gender gap in authorship: a study of 44,000 articles published in 100 high-impact general medical journals. *Eur J Intern Med*. 2022;97:103-105. [doi: [10.1016/j.ejim.2021.09.013](https://doi.org/10.1016/j.ejim.2021.09.013)] [Medline: [34598855](https://pubmed.ncbi.nlm.nih.gov/34598855/)]
16. Filardo G, da Graca B, Sass DM, Pollock BD, Smith EB, Martinez MAM. Trends and comparison of female first authorship in high impact medical journals: observational study (1994-2014). *BMJ*. 2016;352:i847. [FREE Full text] [doi: [10.1136/bmj.i847](https://doi.org/10.1136/bmj.i847)] [Medline: [26935100](https://pubmed.ncbi.nlm.nih.gov/26935100/)]
17. Sebo P, Clair C. Gender inequalities in citations of articles published in high-impact general medical journals: a cross-sectional study. *J Gen Intern Med*. 2023;38(3):661-666. [FREE Full text] [doi: [10.1007/s11606-022-07717-9](https://doi.org/10.1007/s11606-022-07717-9)] [Medline: [35794309](https://pubmed.ncbi.nlm.nih.gov/35794309/)]
18. Gender API. URL: <https://gender-api.com> [accessed 2023-12-31]
19. Sebo P. Performance of gender detection tools: a comparative study of name-to-gender inference services. *J Med Libr Assoc*. 2021;109(3):414-421. [FREE Full text] [doi: [10.5195/jmla.2021.1185](https://doi.org/10.5195/jmla.2021.1185)] [Medline: [34629970](https://pubmed.ncbi.nlm.nih.gov/34629970/)]
20. Sebo P. Use of ChatGPT to explore gender and geographic disparities in scientific peer review. *Open Science Framework*. URL: <https://osf.io/WNRZU/> [accessed 2024-11-29]
21. Sebo P. Gender and geographical inequalities among highly cited researchers: a cross-sectional study (2014-2021). *Intern Emerg Med*. 2023;18(4):1227-1231. [doi: [10.1007/s11739-023-03240-9](https://doi.org/10.1007/s11739-023-03240-9)] [Medline: [36877434](https://pubmed.ncbi.nlm.nih.gov/36877434/)]
22. Negative binomial regression | Stata annotated output. UCLA: Statistical Consulting Group URL: <https://stats.idre.ucla.edu/stata/output/negative-binomial-regression/> [accessed 2023-12-31]
23. Negative binomial regression | Stata data analysis examples. UCLA: Statistical Consulting Group. URL: <https://stats.idre.ucla.edu/stata/dae/negative-binomial-regression/> [accessed 2023-12-31]
24. Klein D. Implementing a general framework for assessing interrater agreement in Stata. *Sage J*. 2018;18(4):871-901. [doi: [10.1177/1536867x1801800408](https://doi.org/10.1177/1536867x1801800408)]
25. Pouris A, Pouris A. The state of science and technology in Africa (2000–2004): a scientometric assessment. *Scientometrics*. 2008;79(2):297-309. [doi: [10.1007/s11192-009-0419-x](https://doi.org/10.1007/s11192-009-0419-x)]
26. Skopec M, Issa H, Reed J, Harris M. The role of geographic bias in knowledge diffusion: a systematic review and narrative synthesis. *Res Integr Peer Rev*. 2020;5:2. [doi: [10.1186/s41073-019-0088-0](https://doi.org/10.1186/s41073-019-0088-0)] [Medline: [31956434](https://pubmed.ncbi.nlm.nih.gov/31956434/)]
27. Lor P. Scholarly publishing and peer review in the Global South: the role of the reviewer. *JLIS.it*. 2022;14(1):10-29. [doi: [10.36253/jlis.it-512](https://doi.org/10.36253/jlis.it-512)]
28. Schneider S, Morgan C, Magill C, Hersh M, Sang K, MacIntosh R. Evidence Review: Peer Review Bias in the Funding Process: Main Themes and Interventions. Edinburgh, Scotland. Heriot-Watt University; 2024.

Abbreviations

AI: artificial intelligence

Edited by A Mavragani; submitted 22.02.24; peer-reviewed by L Zhu, A Hassan, GK Gupta, H Fang; comments to author 09.04.24; revised version received 21.04.24; accepted 12.06.24; published 09.12.24

Please cite as:

Sebo P

Use of ChatGPT to Explore Gender and Geographic Disparities in Scientific Peer Review

J Med Internet Res 2024;26:e57667

URL: <https://www.jmir.org/2024/1/e57667>

doi: [10.2196/57667](https://doi.org/10.2196/57667)

PMID:

©Paul Sebo. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 09.12.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.