Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6, 960-979
2024
Publisher: Learning Gate
DOI: 10.55214/25768484.v8i6.2189

Evident gap between generative artificial intelligence as an academic editor compared to human editors in scientific publishing

Malik Sallam[1,2*], Kholoud Al-Mahzoum[3], Omar Marzoaq[4], Mohammad Alfadhel[3], Amer Al-Ajmi[3], Mansour Al-Ajmi[3], Mohammad Al-Hajeri[3], Muna Barakat[5]
[1]Department of Pathology, Microbiology and Forensic Medicine, School of Medicine, The University of Jordan, Amman 11942, Jordan; malik.sallam@ju.edu.jo (M.S.).
[2]Department of Clinical Laboratories and Forensic Medicine, Jordan University Hospital, Amman 11942, Jordan.
[3]School of Medicine, The University of Jordan, Amman 11942, Jordan.
[4]School of Dentistry, The University of Jordan, Amman 11942, Jordan.
[5]Department of Clinical Pharmacy and Therapeutics, Faculty of Pharmacy, Applied Science Private University, Amman 11931, Jordan.

Abstract: The labyrinthine process of manuscript evaluation in scientific publishing often delays disseminating timely research results. Generative Artificial Intelligence (genAI) models could potentially enhance efficiency in academic publishing. However, it is crucial to scrutinize the reliability of genAI in simulating human editorial decisions. This study analyzed 34 manuscripts authored by the corresponding author, involving initial editorial decisions from six publishers across 28 journals. Two genAI models, ChatGPT-4o and Microsoft Copilot, assessed these manuscripts using tailored prompts. The correlation between genAI and actual human editorial decisions was evaluated using Kendall's τb. The original decision-making speed and the quality of genAI outputs evaluated by the CLEAR tool were recorded. Editorial decision-making by genAI models was instantaneous, compared to the editors' average of 21.6±31.1 days. Both models achieved high scores on the CLEAR tool, averaging 4.8±0.4 for ChatGPT-4o and 4.8±0.5 for Copilot. Despite these efficiencies, there was no significant correlation between the genAI and human decisions (τb=0.121, P=.487 for ChatGPT-4o; τb=0.197, P=.258 for Copilot), nor between the decisions of the two genAI models (τb=0.318, P=.068). This preliminary study indicated that genAI models can expedite the editorial process with high-quality outputs. However, genAI has not yet achieved the accuracy of human editors in decision-making.

Keywords: *AI benchmarking, Editorial policies, Publishing standards.*

1. Introduction

Scholarly communication has traditionally hinged on the processes of manuscript submission, editorial evaluation, and peer review [1,2]. These steps are considered fundamental to scientific discourse, serving to critically appraise and archive scientific knowledge [3,4].

The processes of editorial evaluation and peer review are considered essential to maintain the integrity and quality of scientific literature [2]. However, these critical processes suffer various challenges and inefficiencies that compromise their effectiveness, thereby delaying the timely evaluation and dissemination of scientific knowledge [2]. These issues necessitate a comprehensive analysis and potential innovative reforms to enhance the efficiency and reliability of scientific publishing [5,6].

Recently, Gregory and Dennis outlined the typical workflow of editorial and peer review processes in scientific publishing [7]. Initially, submitted manuscripts undergo preliminary quality assessments, which may include plagiarism checks, depending on the policies of different journals and publishers. Then, the Editor-in-Chief or the assigned academic editor conducts an initial evaluation to determine if

* Correspondence: malik.sallam@ju.edu.jo

the manuscript merits further progress into the peer review. Following peer review, where the manuscript is scrutinized by external experts, the academic editor makes the final decision heavily based on the reviewers' reports [7,8].

The current system of editorial evaluation and peer review exhibits significant weaknesses [9,10]. First, the susceptibility of editorial and peer critical appraisal evaluations to both implicit and explicit biases is a fundamental flaw as highlighted in several studies [10-13]. These studies demonstrated that peer review outcomes can be significantly influenced by possible editors' and reviewers' biases about to the nationality, institutional affiliation, and gender of the authors [11,14–17]. Such biases compromise the objectivity and fairness of the editorial and peer review processes, leading to the marginalization of certain groups within the academic community and negatively affecting the diversity of published research [18,19].

Second, the quality of editorial evaluations and peer reviews can exhibit substantial variability, influenced by factors such as the expertise, workload, and motivation of the reviewers and academic editors [20]. Although many editors and reviewers are recognized for their thoroughness and constructiveness, others may be inattentive and less constructive [7,21]. This inconsistency undermines the reliability and consistency of the editorial and peer review processes, which may result in the publication of flawed research or the unjust rejection of high-quality manuscripts [22,23].

Third, the editorial and peer review systems are increasingly strained by the rapidly growing volume of manuscript submissions [24]. Therefore, the surge in manuscript submissions would predictably lead to prolonged editorial evaluation and review times, placing an increased burden on editors and reviewers and causing further editorial delays [25,26]. The academic editors and reviewers often perform these evaluation duties voluntarily, alongside their own research and professional responsibilities, which exacerbate the strain on the editorial and peer review systems [27,28]. The time-intensive nature of the editorial and review processes is well documented and can significantly delay the dissemination of important research findings, thereby impeding scientific progress [26,29,30].

Fourth, the peer review process commonly operates under double-blinded or single-blinded conditions, where reviews are conducted anonymously. This anonymity can shield reviewers from potential retaliation by authors dissatisfied with critical feedback [31]. However, the blinded conditions also conceal the reviewers' contributions, potentially diminishing the incentive for thorough and thoughtful reviews [32]. Additionally, this lack of transparency in the review process can obscure decision-making, raising concerns about accountability and fairness [33,34].

Finally, academic editors and peer reviewers typically do not receive financial compensation for their significant contributions to manuscript evaluation [35]. Aczel *et al.* characterized this uncompensated effort as 'a billion-dollar donation' highlighting the substantial time and effort editors and reviewers contribute without financial reward [36]. The absence of financial incentives can adversely affect manuscript evaluation process, which is manifested as a limited pool of available editors and reviewers, delays in manuscript evaluation, and editor/reviewer fatigue [37,38].

Thus, the prolonged decision-making process of scientific journals is one of the major challenges in scientific communication and this delay can be pinpointed to certain bottlenecks such as the initial manuscript evaluation by academic editors [39]. In an early study, Azar emphasized that the initial response time, defined as the interval between manuscript submission and the first editorial decision, holds greater importance than other stages of publication delay [40]. Thus, there is a need to investigate innovative methods that could augment or refine the existing editorial evaluation to improve efficiency while maintaining reliability [41]. The emergence of generative artificial intelligence (genAI) promises to enhance this process by addressing these inefficiencies [42-44]. However, challenges in AI-driven editorial decisions are anticipated in terms of possible inaccuracies, bias and indeterminate reliability [42,45].

Thus, this study aimed to assess the correlation between decisions made on real manuscripts by academic editors across various journals and publishers compared to those made by popular genAI

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

models. Specifically, this study aimed to explore the potential of two genAI models, namely ChatGPT-4o and Microsoft Copilot, to serve as academic editors for manuscripts previously evaluated by human editors in diverse scientific journals with varying outcomes.

2. Methods

2.1. Study Design and Inclusion Criteria

This descriptive study was based on the METRICS checklist, which provides a framework for the design and reporting of genAI content evaluation studies [46]. The checklist includes the following criteria: Model, Evaluation, Timing, Range/Randomization, Individual factors, Count, and Specificity of prompts and language [46].

The manuscripts used to evaluate the performance of genAI models in this study were selected based on their prior assessment by human editors across various scientific journals to establish a benchmark for assessing the decision-making capabilities of the genAI models. A key criterion for inclusion was that all selected manuscripts had the first author of this study also serving as the corresponding author. Then, the manuscripts were selected based on the availability of precise data on the number of days from submission to the first editorial decision, which was either to 'reject' or 'send for review'. The selected manuscripts involved two principal research areas of the first author: (1) infectious disease, public health, or vaccination, and (2) AI in healthcare education and practice. Given the original nature of the work and the fields covered, no copyright issues were anticipated. The study was approved by the Institutional Review Board (IRB) at the Faculty of Pharmacy – Applied Science Private University (reference number: 2024-PHA-20) granted on 20 May 2024.

2.2. GenAI Models that Were Selected for Performance Evaluation

For the purpose of this study, two advanced genAI models were utilized: (1) ChatGPT-4o and (2) Microsoft Copilot. These models were selected based on their state-of-the-art capabilities in natural language processing and their prevalent use in academic settings. To maintain the consistency of the results, specific features of the genAI models were intentionally not utilized. The 'regenerate response' and 'modify response' functions were not used during genAI prompting to eliminate potential biases that could arise from iterative refinements or human-influenced adjustments.

2.3. Evaluation of genAI Models' Output

The evaluation of the genAI-generated content was conducted using a two-pronged approach. First, the content evaluation was independently performed by the first and senior authors of this study, both of whom are experienced academic editors across various scientific journals. A modified version of the CLEAR tool was employed to assess the genAI-generated content [47], which evaluates quality across three dimensions: (1) completeness of the response, (2) accuracy reflecting the absence of incorrect information and evidence-based content, and (3) appropriateness and relevance, focusing on clarity, organization, and relevance of the generated content [47]. Each dimension was rated on a 5-point Likert scale, ranging from 1 (poor) to 5 (excellent). Second, to measure the genAI models' performance, the decisions made by the genAI models were compared against the original decisions recorded by human editors—specifically, whether manuscripts were 'sent for review' or 'rejected'.

2.4. Timing of genAI Models' Prompting

To ensure consistency and control in the genAI testing conditions, the evaluation of the two models was scheduled precisely over a defined period, from August 27 to August 30, 2024. This narrow window of time aimed to minimize external variations that could affect the performance and outcomes of the genAI evaluations including model updates.

2.5. Range of Manuscripts' Topics and Randomization of Manuscript Selection

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

The study involved manuscripts that covered a broad range of topics as defined by Scopus SciVal [48]. These topics included: (1) Artificial Intelligence; Diagnostic Imaging; Radiology, (2) Artificial Intelligence; Engineering Education; Machine Learning, (3) Conspiracy Theory; Public Health; COVID-19, (4) *Coronavirinae*; SARS Coronavirus; COVID-19, (5) *Entamoeba* Gingivalis; Periodontitis; *Trichomonas*, (6) Hepatitis C; Antivirus Agent; Pharmacotherapy, (7) Human Respiratory Syncytial Virus; Palivizumab; Neonatal Infant, (8) Papillomavirus Infection; Wart Virus Vaccine; Adolescent, (9) Papillomavirus Infection; Wart Virus Vaccine; Adolescent, (10) Support Vector Machine; Cognitive Level; Machine Learning, (11) Vaccine Efficacy; Vaccination Policy; COVID-19, (12) Viral Pneumonia; Coronavirinae; COVID-19.

Data on the time from manuscript submission to the first editorial decision in days and details about the journals' publishers were recorded as well based on the different journals submission systems and e-mail communications with the corresponding author. Due to the restricted availability of manuscripts with complete data from submission to the initial editorial decision, randomization was not feasible.

2.6. Individual Factors in Prompting and Evaluation of Genai Content

To ensure uniformity in the evaluation of genAI-generated content, all interactions with the two genAI models were conducted by the first author using the same electronic device. The CLEAR tool was employed to systematically assess the quality of genAI content [47]; however, this tool depended on the subjective evaluations by the first and senior authors. The inherent subjectivity in the human editorial decisions, which served as benchmarks for comparing genAI performance, is recognized as another subjective factor in the evaluation process.

2.7. Minimum Count of Manuscripts Needed for Evaluation

The sample size for this study was determined to ensure adequate statistical power for detecting a significant correlation using Kendall's τb, a non-parametric measure ideal for assessing the association between two sets of ranked data. This analysis aimed to capture an expected moderate positive relationship. To maintain a high level of precision, we cited Bujang guidelines for Kendall's τb that recommend a minimum of 30 manuscripts to achieve a target correlation with a 95% confidence level [49].

2.8. Specificity of Prompts and Language Used

The genAI models were provided with a standardized prompt to ensure consistency in the evaluation process. Each genAI model was presented with the title and abstract of a manuscript, followed by the following specific prompt tailored to the aims and scope of the respective journal: "Please act as the editor-in-chief for the scientific journal "X". Based on the aims and scope detailed on the journal's homepage, you will evaluate the provided manuscript title and abstract. Your decision should be informed by the ICMJE/COPE guidelines, focusing on the manuscript's relevance, novelty, clarity, methodological rigor, and ethical considerations. You must choose between "send for peer review" or "reject submission" and provide a scientific and ethical rationale for your decision in no more than 250 words." The prompt was delivered in English, ensuring that the genAI models' responses were directly comparable across all cases, maintaining the uniformity of the evaluation process.

2.9. Statistical and Data Analysis

Statistical analyses were conducted using IBM SPSS Statistics Version 26.0 (Armonk, NY: IBM Corp). Associations between categorical variables were examined using the two-sided Fisher's Exact test (FET), which provides robust significance testing for small datasets.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

The CLEAR scores, treated as scale variables, were assessed for relationships with categorical variables using the Mann–Whitney *U* and the Kruskal-Wallis tests. These non-parametric tests are suitable for data distributions that deviate from normality, a condition confirmed for the CLEAR score distribution via the Kolmogorov-Smirnov test ($P<.001$).

To assess the reliability of inter-rater evaluations, the Intraclass Correlation Coefficient (ICC) was used. In addition, one-sided Analysis of Variance (ANOVA) was utilized to examine the variability across components of CLEAR, assessing differences within the grouped variables.

Correlations between AI-generated evaluations and decisions made by human editors were analyzed using Kendall's τb, an appropriate non-parametric measure for ranked data as well as via calculation of the Cohen's κ. We established a significance level of 0.05 to determine statistical significance across all analyses.

Manuscripts were categorized into two thematic groups for further analysis: (1) Infectious disease, public health, or vaccination, and (2) AI in healthcare education and practice.

Diagnostic performance metrics such as sensitivity, specificity, positive predictive values (PPVs), negative predictive values (NPVs), and Receiver Operating Characteristic (ROC) curves were also calculated to evaluate the accuracy of AI decision-making relative to human editors.

3. Results

3.1. Features of the Selected Manuscripts and Human Editorial Decisions

A total of 34 manuscripts with first editorial decisions by 28 different scientific journals representing six different publishers were used. Twenty-one manuscripts were rejected by the editors (61.8%), while 13 were sent for peer review (38.2%). The overall time to first decision for the entire sample of manuscripts was 21.6±31.1 days. Simple description of the duration of editorial decision in days, publishers, and manuscript topics is shown in Table 1.

Table 1.
Description of the manuscripts included as the final dataset ($N=34$).

Variable	Category	Original academic editor decision	
		Send for review	Reject
Duration of original decision in days	Mean±SD	24.3±33.1	19.9±30.5
Publisher	Dove medical press	2 (100)	0 (0)
	Frontiers media S.A.	4 (100)	0 (0)
	Multidisciplinary digital publishing institute	0 (0)	11 (100)
	Oxford university press	0 (0)	1 (100)
	Springer nature/Springer publishing company	4 (36.4)	7 (63.6)
	Taylor & francis	3 (60.0)	2 (40.0)
Manuscript topic	Infectious disease, public health, or vaccination	7 (43.8)	9 (56.3)
	AI in healthcare education and practice	6 (33.3)	12 (66.7)

Note: AI: Artificial intelligence; SD: Standard deviation.

Publishers with more than two manuscripts were evaluated to assess if there is a difference in time to first editorial decision regardless of the decision. Four publishers were assessed as shown in Figure 1, with the fastest time to first editorial decision observed with Multidisciplinary Digital Publishing

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

Institute (MDPI) (mean: 4.1±1.8 days) followed by Frontiers (mean: 4.5±1.3 days), Taylor & Francis (mean: 13.2±14.1 days), while the slowest time was observed for Springer (mean: 4.1±1.8 days, *P*=.001).

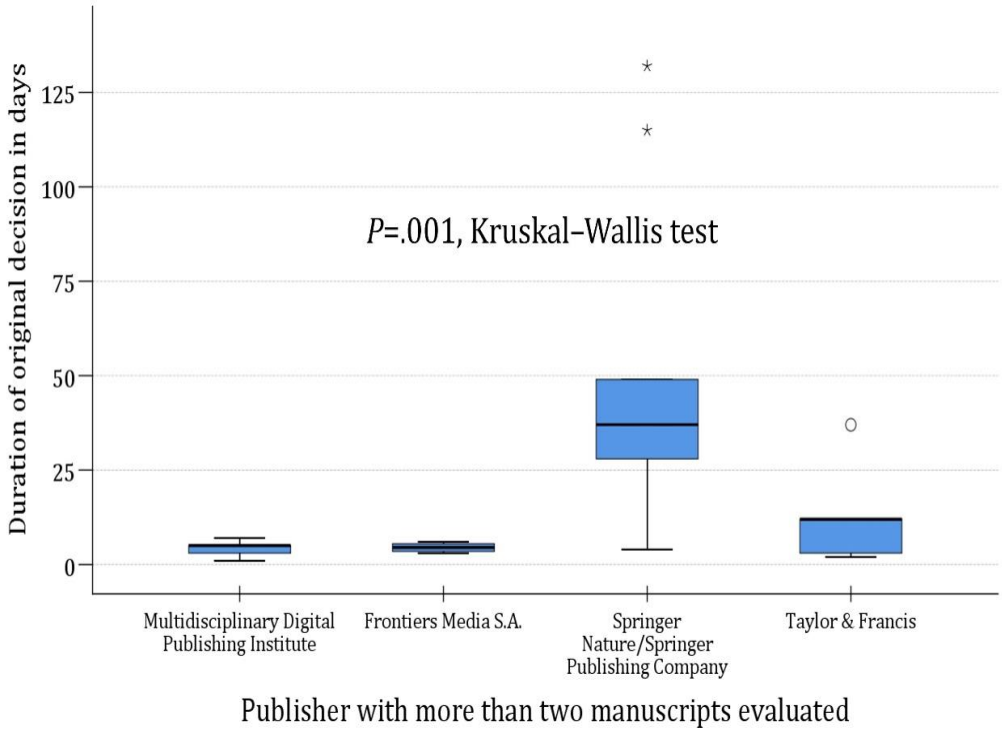

Figure 1.
Whisker plots for the distribution of time to first editorial decision for the manuscripts submitted to different publishers.

3.2. GenAI Decisions And Output Based On the CLEAR Scores

For the genAI decisions, ChatGPT-4o recommended rejection of 13 manuscripts (38.2%) compared to only two rejections by Copilot (5.9%, *P*=.139, FET).

Stratified analyses revealed that ChatGPT-4o recommended sending 81.3% of manuscripts related to infectious diseases, public health, or vaccination for review and rejected 18.8% compared to rejection of a majority (55.6%) of the manuscripts in the topic of AI in healthcare education and practice (*P*=.039). Additionally, manuscripts that ChatGPT-4o rejected had a significantly longer original decision time of 33.6±36.4 days compared to 14.1±25.4 days for those it recommended for review (*P*=.022). Copilot showed no statistically significant differences in decision-making patterns based on manuscript topic, or original decision duration as shown in Table 2.

Table 2.
Generative AI models' editorial decisions.

Variable	Category	ChatGPT-4o decision	*P* value	Copilot decision	*P* value

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

		Send for review	Reject		Send for review	Reject	
		Count (%)	Count (%)		Count (%)	Count (%)	
Original academic editor decision	Send for review	9 (69.2)	4 (30.8)	0.718	13 (100)	0 (0)	0.513
	Reject	12 (57.1)	9 (42.9)		19 (90.5)	2 (9.5)	
Manuscript topic	Infectious disease, public health, or vaccination	13 (81.3)	3 (18.8)	0.039	16 (100)	0 (0)	0.487
	AI in healthcare education and practice	8 (44.4)	10 (55.6)		16 (88.9)	2 (11.1)	
		Mean±SD	Mean±SD		Mean±SD	Mean±SD	
Average ChatGPT-4o clear score		4.9±0.2	4.6±0.6	0.232			
Average Copilot clear score					4.7±0.5	5±0	0.556
Duration of original decision in days		14.1±25.4	33.6±36.4	0.022	22.4±31.9	8.5±4.9	0.970

Note: AI: Artificial intelligence; SD: Standard deviation; *P* values were calculated using the two-sided Fisher's exact test and the Mann Whiteny *U* test.

For the evaluation of content generated by both genAI models acting as academic editors, The CLEAR scores showed no statistically significant differences between manuscripts recommended for review and those rejected by either ChatGPT-4o or Copilot, with both models consistently achieving average scores that were rated as excellent, reflecting high overall quality in their content evaluations.

Specifically, in ChatGPT-4o, zero variance was observed for the completeness dimension with identical rating of 5.0 by both raters; for accuracy, the average was 4.66±0.57, and for the relevance the average was 4.74±0.58 (*P*=.001, one way ANOVA). For Copilot, the completeness showed an average of 4.91±0.33; the average was 4.63±0.78 for accuracy and for relevance the average was 4.75±0.53 (*P*=.021, one way ANOVA).

To assess the reliability of these evaluations, the ICC was used. For ChatGPT-4o, completeness had no variance, consistently receiving the maximum score of 5.0. For accuracy, the average measures ICC was 0.959 (95% CI: 0.919 – 0.980), indicating high reliability. For relevance, the average measures ICC was 0.906 (95% CI: 0.811 – 0.953), also reflecting strong agreement. Similarly, Copilot's assessments showed an average measures ICC of 0.845 (95% CI: 0.691 – 0.923) for completeness, 0.988 (95% CI: 0.976 – 0.994) for accuracy, and 0.847 (95% CI: 0.694 – 0.924) for relevance, all of which indicate high reliability in the Copilot content evaluations by the two raters.

3.3. Correlation Between genAI and Human Editorial Decisions

The analysis of correlations between genAI models and human editorial decisions revealed no significant correlation. For ChatGPT-4o, the Cohen κ was 0.108 while the Kendall's τb correlation coefficient was 0.121 (*P*=.487), indicating a statistically insignificant association between the decisions made by ChatGPT-4o and those made by human editors. Similarly, for Copilot, the Cohen κ was 0.074 and the Kendall's τb correlation coefficient was 0.197 (*P*=.258), also reflecting no significant relationship between Copilot and human decisions. Furthermore, when comparing the decisions between the two genAI models themselves, the Cohen κ was 0.183 and the Kendall's τb was 0.318 (*P*=.068), suggesting a trend toward correlation but failing to reach statistical significance.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

In evaluating the performance of ChatGPT-4o against the original academic editor decisions, the sensitivity of ChatGPT-4o was found to be 69.2%, indicating that it correctly identified 69.2% of the manuscripts that the original editors sent for review. The specificity was 42.9%, reflecting a lower ability to correctly identify manuscripts that were rejected by the original editors, despite providing excellent content for the reasons behind rejection (Supplementary Table). The PPV was also 42.9%, meaning that when ChatGPT-4o recommended sending a manuscript for review, it matched the original editor's decision 42.9% of the time. Conversely, the NPV was 69.2%, indicating that when ChatGPT-4o recommended rejecting a manuscript, it correctly aligned with the original editor's decision in 69.2% of cases. These results suggest a moderate alignment between ChatGPT-4o decisions and those of the human editors, particularly in identifying manuscripts suitable for review.

The diagnostic performance of Copilot in predicting the final decision outcome was assessed against the original academic editor decisions as well. Copilot demonstrated a sensitivity of 100%, correctly identifying all manuscripts that the original editors sent for review. However, its specificity was only 9.5%, indicating a low ability to correctly reject manuscripts that were also rejected by the original editors. The PPV was 40.6%, meaning that when Copilot recommended sending a manuscript for review, it matched the original editor decision 40.6% of the time. The NPV was 100%, showing that when Copilot recommended rejecting a manuscript, it was fully aligned with the original editor decision (Figure 2. These results highlight that while Copilot is highly sensitive in identifying manuscripts suitable for review, it has a lower specificity, often recommending review for manuscripts that the original editors rejected.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

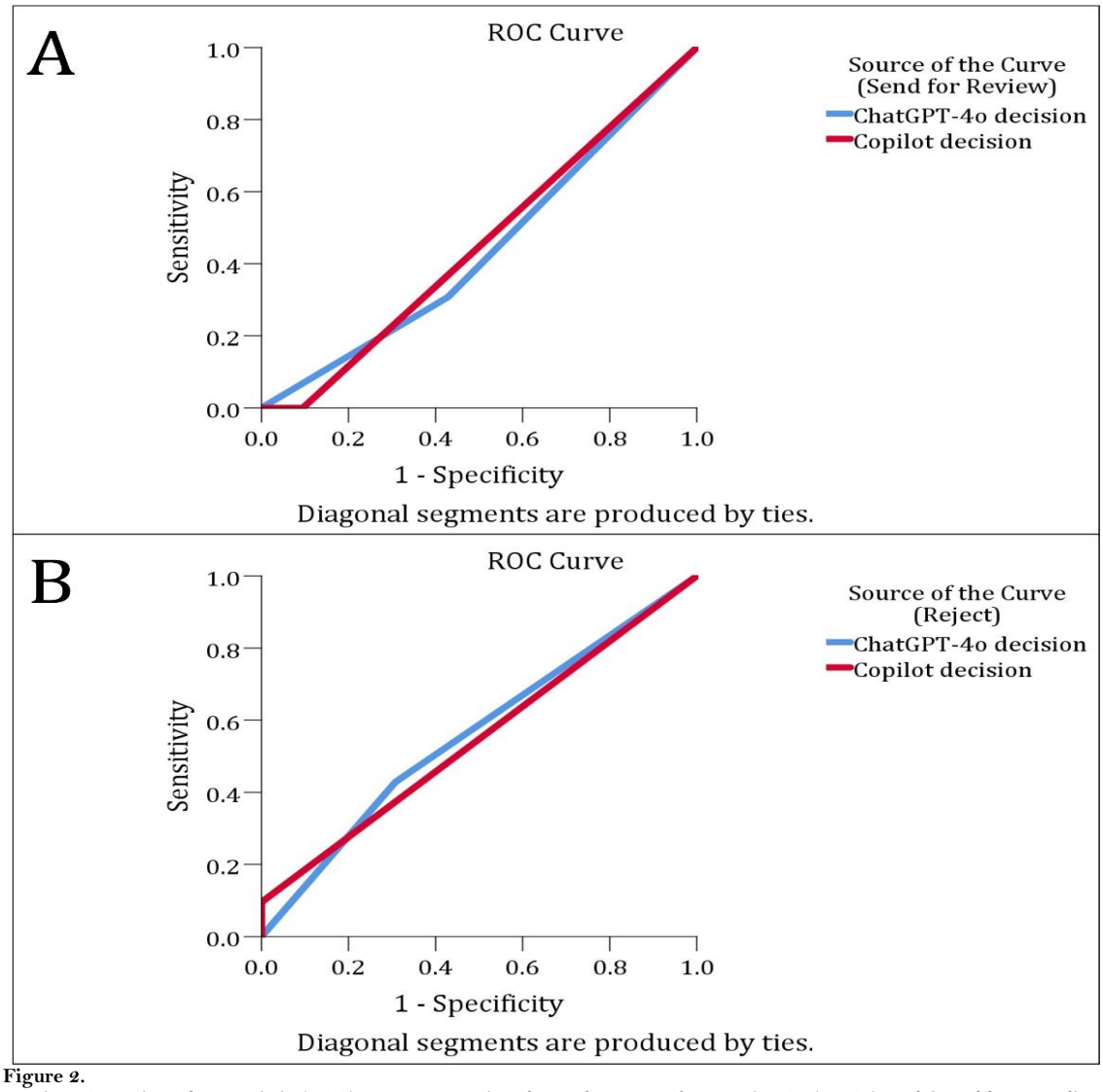

Figure 2.
Receiver Operating Characteristic (ROC) curves comparing the performance of generative AI (genAI) models and human editors in decision-making. Panel (A) illustrates the ROC curve for human decisions "Send for Review" while panel (B) depicts the curve for "Reject" decisions.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

4. Discussion

The integration of genAI models into the editorial processes of academic journals holds promising perspectives to enhance manuscript evaluation efficiency. However, as demonstrated by the findings of this study, significant current challenges remain which would hinder the effective implementation of such genAI technologies. The discordance between decisions made by human editors and those made by genAI models (ChatGPT-4o and Copilot) highlighted a key limitation. This limitation was manifested in the low performance measures shown by the two models in reaching similar decisions compared to human editors.

Despite the challenges manifested as difficulty of genAI models' ability to replicate the human editor decisions, the potential benefits of genAI integration into editorial practices cannot be overlooked. The ability of genAI to provide rapid feedback could revolutionize the publishing process by significantly reducing the time from submission to decision—a critical factor given the long publication times observed in recent studies. For example, Manganaro reported an alarming trend in biomedical research, where publication times have extended by an average of four weeks between 2014 and 2019 [50]. Similarly, Luwel *et al.* highlighted that despite reductions in editorial delays for papers in Mathematics & Computer Science and Social Sciences & Humanities, these periods remain considerably longer than the median across all disciplines [51]. Collectively, these observations besides our finding of the delay in reaching an initial editorial decision highlighted the need for innovative solutions to enhance the publication process. The swift feedback capability of genAI offers a promising solution to address these delays. However, to realize this genAI potential in academic publishing, significant improvements are necessary to align genAI's decision-making processes more closely with the complex judgment criteria employed by human editors.

The study findings highlighted a key challenge in the integration of genAI into scientific publishing manifested in the discordance between decisions made by human editors and those generated by genAI models. This discordance can be attributed to the context-sensitive nature of editorial decision-making, which involves evaluating factors like novelty, relevance, and journal scope beyond the mere scientific merit as illustrated by Gilliland & Cortina [52]. Despite the ability of the current generation of genAI models to process and synthesize intellectually plausible content, there is evidence showing caveats in genAI content including the lack of depth to fully replicate complex human judgments [53,54].

The lack of significant correlation between AI-generated and human decisions emphasizes the need for caution in employing genAI in editorial processes. The study findings also emphasize the necessity for developing more sophisticated genAI models, tailored to mirror human editorial practices. Future improvements may benefit from training genAI models on historical editorial decisions with robust feedback and benchmarking mechanisms [55]. In turn, this would enhance genAI's understanding of editorial criteria, better aligning its capabilities with the intricate demands of scholarly publishing.

Conversely, the objectivity offered by genAI could potentially enhance author satisfaction through objective and prompt editorial decision-making aligning with previous evidence demonstrating the editorial impact on the quality and quantity of a journal's publications [56]. Key among these is the ability of genAI to generate instantaneous decisions, dramatically reducing the average three-week wait associated with human editors as found in this study. This rapid decision-making could greatly accelerate the dissemination of knowledge, a particular advantage during times of swift scientific evolution or public health crises without compromising the quality of publications [57]. Additionally, genAI models have the potential to provide constructive helpful feedback compared to the often cliché and unhelpful rejection letters that authors usually receive, which cite general issues like lack of novelty or fit with the journal's scope. Additionally, genAI has the potential to enhance peer review by providing more detailed critiques and examples, addressing the issue of overly harsh comments from human reviewers as shown in the study by Hyland & Jian [21].

This enhanced feedback mechanism aligns with findings from Huisman & Smits, who noted that efficiency in editorial processes correlates strongly with higher author satisfaction—especially in fields

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

where peer reviews are typically protracted [58]. Authors in such domains anticipate and value in-depth, informative interactions with editors, an expectation that genAI could potentially fulfill. Furthermore, assessments made in this study using the CLEAR scores, which evaluate content quality, indicated that the quality of evaluations by genAI models matches that of human editors, suggesting that any discrepancies in decision-making may not stem from deficiencies in genAI evaluations but rather from differing interpretations of a manuscript's merits.

However, while some CLEAR score components like completeness were consistently high in this study, others, such as accuracy and relevance, showed a room for improvement. These subtle but critical differences underline the need for ongoing refinement of genAI training datasets and algorithms to better mimic the sophisticated decision-making processes of human editors, thereby enhancing the overall efficacy and reliability of genAI in academic publishing [59]. This result was different compared to recent studies where the aspects such as completeness, relevance, and accuracy vary based on topic, depth of cognitive abilities needed, and language used [54,60,61]. Thus, the variable performance of genAI models should be considered if genAI is to be implemented in scientific publishing.

The study results also revealed distinct decision-making patterns between ChatGPT-4o and Copilot in handling academic manuscripts. ChatGPT-4o exhibited a conservative approach, particularly in AI-related topics, suggesting a high sensitivity to the complexities inherent in AI research. The cautious approach of ChatGPT-4o, similar to that of a specialized human editor, emphasizes precision and rigor, possibly to avoid premature acceptance of flawed research (avoidance of type I errors). In contrast, Copilot adopted a more permissive, less discerning approach, which aligns with a generalist editorial style, focusing on rapid dissemination of findings, potentially at the expense of detail and rigor, reflecting a tendency to minimize missed opportunities for important discoveries (type II errors). Taken together, this variability in genAI model performance suggests the need to consider the AI models different focuses and potential areas for refinement to better align with the complex needs of academic publishing.

For stakeholders in the academic publishing industry, including authors, editors, peer reviewers, and large publishing companies, the current study findings highlight the potential and limitations of integrating genAI into editorial processes. AI has the capability to expedite decision-making significantly; however, current limitations in the accuracy of genAI judgment necessitate a balanced approach where AI supports rather than supplants human decision-makers [62]. This study showed that the use of genAI could streamline the decision-making process, where typical delays span from a few weeks to several months, depending on the publisher. Improving the accuracy of genAI could address these delays effectively.

Finally, it is worth mentioning that this study incidentally found that all 11 manuscripts submitted to various MDPI journals were rejected. Notably, manuscripts sent for peer review at MDPI, which typically receive initial decisions within a few days, were excluded from the analyzed dataset due to the inability to track the exact duration to the first editorial decision. The frequent submissions to MDPI by the first author were motivated by MDPI journals' broad scope, inclusion of many journals in major indexing databases like Scopus, Web of Science, and PubMed, and notably, the MDPI rapid editorial decisions and efficient peer review [63,64]. The prompt responses from assistant editors and clear communication with guest editors also encouraged submissions. These observations emphasize the need to consider the broader editorial practices at MDPI and necessitate further detailed investigation into their editorial practices without prematurely attributing the publisher's growth to criticisms of their operations [65].

It is important to note that interpreting the current study should be done with extreme caution due to the following limitations. First, the modest sample size and limited diversity in manuscript topics and publishers may restrict the generalizability of findings across various scientific fields. This was compounded by the inherent homogeneity of the sample, as all the included manuscripts were co-authored by the first author. Second, the genAI models used in the study could harbor inherent biases

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

due to training datasets not fully representing the diverse decision-making criteria across scientific journals. Third, the content evaluation metrics in terms of the CLEAR scores may not capture all qualitative aspects of editorial quality such as complex arguments or theoretical significance. Fourth, the binary nature of sensitivity and specificity metrics oversimplifies the complex cognitive processes involved in human editorial decisions, which include considerations of impact, reader interest, and novelty. Fifth, the current study assumed that the decisions by the human editors were the "correct" decisions; however, this assumption lacks empirical validation and may overlook potential biases or inconsistencies in editorial judgments. Finally, there is a possibility that using a different prompt could alter the editorial decisions and this should be considered in future research.

5. Conclusions

The study showed that the integration of genAI into editorial processes in scientific publishing currently faces substantial challenges, primarily in aligning the speed of automated decisions with the depth required in human editorial judgment. However, genAI models have a significant potential to enhance the efficiency of editorial processes in scientific publishing. This study underscored the need for ongoing refinement of genAI models to better mimic complex human decision-making, ensuring the integrity of scholarly communication. As genAI technologies evolve, they promise to not only accelerate publication processes but also improve the quality and responsiveness of editorial decisions, thereby benefiting the broader academic community.

Authorship:

Conceptualization: Malik Sallam; **Data curation:** Malik Sallam; **Formal analysis:** Malik Sallam, Kholoud Al-Mahzoum, Omar Marzoaq, Amer Al-Ajmi, Mansour Al-Ajmi, Mohammad Al-Hajeri, Mohammad Alfadhel, and Muna Barakat; **Investigation:** Malik Sallam, Kholoud Al-Mahzoum, Omar Marzoaq, Amer Al-Ajmi, Mansour Al-Ajmi, Mohammad Al-Hajeri, Mohammad Alfadhel, and Muna Barakat; **Methodology:** Malik Sallam and Muna Barakat; **Project administration:** Malik Sallam; **Software:** Malik Sallam; **Resources:** Malik Sallam and Muna Barakat; **Supervision:** Malik Sallam; **Validation:** Malik Sallam and Muna Barakat; **Visualization:** Malik Sallam; **Writing – original draft:** Malik Sallam; **Writing – review & editing:** Malik Sallam, Kholoud Al-Mahzoum, Omar Marzoaq, Amer Al-Ajmi, Mansour Al-Ajmi, Mohammad Al-Hajeri, Mohammad Alfadhel, and Muna Barakat.

Data Accessibility Statement:

The datasets analyzed during the current study are available in the public data tool Open Science Framework (OSF), using the following link: https://osf.io/rvq4f/; DOI: 10.17605/OSF.IO/RVQ4F

Ethical Approval:

The study was approved by the Institutional Review Board (IRB) at the Faculty of Pharmacy – Applied Science Private University (reference number: 2024-PHA-20) granted on 20 May 2024.

Competing Interests:

The authors have declared no competing interests. Malik Sallam is an academic editor in PLOS ONE and Associate Editor for Virology in Frontiers in Microbiology, and Associate Editor for Vaccine Safety and Regulation in Frontiers in Drug Safety and Regulation. Muna Barakat is the Editor-in-Chief for the Jordan Journal of Applied Science - Natural Science Series.

References

[1] Maron, N.; Smith, K. Current Models of Digital Scholarly Communication: Results of an Investigation Conducted by Ithaka Strategic Services for the Association of Research Libraries. *The Journal of Electronic Publishing* **2009**, *12*, doi:10.3998/3336451.0012.105.

[2] Banks, D. Thoughts on Publishing the Research Article over the Centuries. *Publications* **2018**, *6*, 10, doi:10.3390/publications6010010.

[3] Tumin, D.; Tobias, J.D. The peer review process. *Saudi J Anaesth* **2019**, *13*, S52-s58, doi:10.4103/sja.SJA_544_18.

[4] Kelly, J.; Sadeghieh, T.; Adeli, K. Peer Review in Scientific Publications: Benefits, Critiques, & A Survival Guide. *Ejifcc* **2014**, *25*, 227-243.

[5] Horbach, S.P.J.M.; Halffman, W. The changing forms and expectations of peer review. *Research Integrity and Peer Review* **2018**, *3*, 8, doi:10.1186/s41073-018-0051-5.

[6] Hall, R.P., 3rd. Effective Peer Review: Who, Where, or What? *JID Innov* **2022**, *2*, 100162, doi:10.1016/j.xjidi.2022.100162.

[7] Gregory, A.T.; Denniss, A.R. Everything You Need to Know About Peer Review — The Good, The Bad and The Ugly. *Heart, Lung and Circulation* **2019**, *28*, 1148-1153, doi:10.1016/j.hlc.2019.05.171.

[8] Card, D.; DellaVigna, S. What Do Editors Maximize? Evidence from Four Economics Journals. *The Review of Economics and Statistics* **2020**, *102*, 195-217, doi:10.1162/rest_a_00839.

[9] Peh, W.C. Peer review: concepts, variants and controversies. *Singapore Med J* **2022**, *63*, 55-60, doi:10.11622/smedj.2021139.

[10] Haffar, S.; Bazerbachi, F.; Murad, M.H. Peer Review Bias: A Critical Review. *Mayo Clinic Proceedings* **2019**, *94*, doi:10.1016/j.mayocp.2018.09.004.

[11] Lee, C.J.; Sugimoto, C.R.; Zhang, G.; Cronin, B. Bias in peer review. *Journal of the American Society for Information Science and Technology* **2013**, *64*, 2-17, doi:10.1002/asi.22784.

[12] Le Sueur, H.; Dagliati, A.; Buchan, I.; Whetton, A.D.; Martin, G.P.; Dornan, T.; Geifman, N. Pride and prejudice - What can we learn from peer review? *Med Teach* **2020**, *42*, 1012-1018, doi:10.1080/0142159x.2020.1774527.

[13] Smith, R. Classical peer review: an empty gun. *Breast Cancer Res* **2010**, *12 Suppl 4*, S13, doi:10.1186/bcr2742.

[14] Kaatz, A.; Gutierrez, B.; Carnes, M. Threats to objectivity in peer review: the case of gender. *Trends Pharmacol Sci* **2014**, *35*, 371-373, doi:10.1016/j.tips.2014.06.005.

[15] Ben Messaoud, K.; Schroter, S.; Richards, M.; Gayet-Ageron, A. Analysis of peer reviewers' response to invitations by gender and geographical region: cohort study of manuscripts reviewed at 21 biomedical journals before and during covid-19 pandemic. *Bmj* **2023**, *381*, e075719, doi:10.1136/bmj-2023-075719.

[16] Squazzoni, F.; Bravo, G.; Farjam, M.; Marusic, A.; Mehmani, B.; Willis, M.; Birukou, A.; Dondio, P.; Grimaldo, F. Peer review and gender bias: A study on 145 scholarly journals. *Sci Adv* **2021**, *7*, doi:10.1126/sciadv.abd0299.

[17] Helmer, M.; Schottdorf, M.; Neef, A.; Battaglia, D. Gender bias in scholarly peer review. *Elife* **2017**, *6*, doi:10.7554/eLife.21718.

[18] Strauss, D.; Gran-Ruaz, S.; Osman, M.; Williams, M.T.; Faber, S.C. Racism and censorship in the editorial and peer review process. *Frontiers in Psychology* **2023**, *14*, doi:10.3389/fpsyg.2023.1120938.

[19] Bancroft, S.F.; Ryoo, K.; Miles, M. Promoting equity in the peer review process of journal publication. *Science Education* **2022**, *106*, 1232-1248, doi:10.1002/sce.21733.

[20] Tennant, J.P.; Ross-Hellauer, T. The limitations to our understanding of peer review. *Research Integrity and Peer Review* **2020**, *5*, 6, doi:10.1186/s41073-020-00092-1.

[21] Hyland, K.; Jiang, F. "This work is antithetical to the spirit of research": An anatomy of harsh peer reviews. *Journal of English for Academic Purposes* **2020**, *46*, 100867, doi:10.1016/j.jeap.2020.100867.

[22] Smith, R. Peer review: a flawed process at the heart of science and journals. *J R Soc Med* **2006**, *99*, 178-182, doi:10.1177/014107680609900414.

[23] Strauss, D.; Gran-Ruaz, S.; Osman, M.; Williams, M.T.; Faber, S.C. Racism and censorship in the editorial and peer review process. *Front Psychol* **2023**, *14*, 1120938, doi:10.3389/fpsyg.2023.1120938.

[24] Landhuis, E. Scientific literature: Information overload. *Nature* **2016**, *535*, 457-458, doi:10.1038/nj7612-457a.

[25] Teixeira da Silva, J.A.; Dobránszki, J. Excessively Long Editorial Decisions and Excessively Long Publication Times by Journals: Causes, Risks, Consequences, and Proposed Solutions. *Publishing Research Quarterly* **2017**, *33*, 101-108, doi:10.1007/s12109-016-9489-9.

[26] Zhang, Y.; Wang, Y. Understanding delays in publishing interdisciplinary research. *Information Processing & Management* **2024**, *61*, 103826, doi:10.1016/j.ipm.2024.103826.

[27] Nguyen, V.M.; Haddaway, N.R.; Gutowsky, L.F.; Wilson, A.D.; Gallagher, A.J.; Donaldson, M.R.; Hammerschlag, N.; Cooke, S.J. How long is too long in contemporary peer review? Perspectives from authors publishing in conservation biology journals. *PLoS One* **2015**, *10*, e0132557, doi:10.1371/journal.pone.0132557.

[28] Hyland, K. Peer review: Objective screening or wishful thinking? *Journal of English for Research Publication Purposes* **2020**, *1*, 51-65, doi:10.1075/jerpp.19010.hyl.

[29] Björk, B.-C.; Solomon, D. The publishing delay in scholarly peer-reviewed journals. *Journal of Informetrics* **2013**, *7*, 914-923, doi:10.1016/j.joi.2013.09.001.

[30] Xie, Y.; Wu, Q.; Wang, Y.; Hou, L.; Liu, Y. Does the handling time of scientific papers relate to their academic impact and social attention? Evidence from Nature, Science, and PNAS. *Journal of Informetrics* **2024**, *18*, 101504, doi:10.1016/j.joi.2024.101504.

[31] Comer, D.R.; Schwartz, M. The problem of humiliation in peer review. *Ethics and Education* **2014**, *9*, 141-156, doi:10.1080/17449642.2014.913341.

[32] Shoham, N.; Pitman, A. Open versus blind peer review: is anonymity better than transparency? *BJPsych Advances* **2020**, *27*, 1-8, doi:10.1192/bja.2020.61.

[33] Besançon, L.; Rönnberg, N.; Löwgren, J.; Tennant, J.P.; Cooper, M. Open up: a survey on open and non-anonymized peer reviewing. *Res Integr Peer Rev* **2020**, *5*, 8, doi:10.1186/s41073-020-00094-z.

[34] Hope, A.A.; Munro, C.L. Criticism and Judgment: A Critical Look at Scientific Peer Review. *Am J Crit Care* **2019**, *28*, 242-245, doi:10.4037/ajcc2019152.

[35] Warne, V. Rewarding reviewers – sense or sensibility? A Wiley study explained. *Learned Publishing* **2016**, *29*, 41-50, doi:10.1002/leap.1002.

[36] Aczel, B.; Szaszi, B.; Holcombe, A.O. A billion-dollar donation: estimating the cost of researchers' time spent on peer review. *Res Integr Peer Rev* **2021**, *6*, 14, doi:10.1186/s41073-021-00118-2.

[37] Breuning, M.; Backstrom, J.; Brannon, J.; Gross, B.I.; Widmeier, M. Reviewer Fatigue? Why Scholars Decline to Review their Peers' Work. *PS: Political Science & Politics* **2015**, *48*, 595-600, doi:10.1017/S1049096515000827.

[38] Gurwitz, D. Peer review: Award bonus points to motivate reviewers. *Nature* **2017**, *542*, 414, doi:10.1038/542414d.

[39] Amat, C.B. Editorial and publication delay of papers submitted to 14 selected Food Research journals. Influence of online posting. *Scientometrics* **2008**, *74*, 379-389, doi:10.1007/s11192-007-1823-8.

[40] Azar, O.H. Rejections and the importance of first response times. *International Journal of Social Economics* **2004**, *31*, 259-274, doi:10.1108/03068290410518247.

[41] Horbach, S.P.J.M.; Halffman, W. Journal Peer Review and Editorial Evaluation: Cautious Innovator or Sleepy Giant? *Minerva* **2020**, *58*, 139-161, doi:10.1007/s11024-019-09388-z.

[42] Sallam, M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare (Basel)* **2023**, *11*, 887, doi:10.3390/healthcare11060887.

[43] Kuznetsov, I.; Afzal, O.M.; Dercksen, K.; Dycke, N.; Goldberg, A.; Hope, T.; Hovy, D.; Kummerfeld, J.K.; Lauscher, A.; Leyton-Brown, K. What Can Natural Language Processing Do for Peer Review? *arXiv preprint arXiv:2405.06563* **2024**.

[44] Sallam, M.; Al-Farajat, A.; Egger, J. Envisioning the Future of ChatGPT in Healthcare: Insights and Recommendations from a Systematic Identification of Influential Research and a Call for Papers. *Jordan Medical Journal* **2024**, *58*, 95-108, doi:10.35516/jmj.v58i1.2285.

[45] Borji, A. *A Categorical Archive of ChatGPT Failures*; 2023.

[46] Sallam, M.; Barakat, M.; Sallam, M. A Preliminary Checklist (METRICS) to Standardize the Design and Reporting of Studies on Generative Artificial Intelligence-Based Models in Health Care Education and Practice: Development Study Involving a Literature Review. *Interact J Med Res* **2024**, *13*, e54704, doi:10.2196/54704.

[47] Sallam, M.; Barakat, M.; Sallam, M. Pilot Testing of a Tool to Standardize the Assessment of the Quality of Health Information Generated by Artificial Intelligence-Based Models. *Cureus* **2023**, *15*, e49373, doi:10.7759/cureus.49373.

[48] Elsevier. SciVal 10. Available online: https://www.elsevier.com/products/scival (accessed on 4 September 2024).

[49] Bujang, M.A. An elaboration on sample size determination for correlations based on effect sizes and confidence interval width: a guide for researchers. *Restor Dent Endod* **2024**, *49*, e21, doi:10.5395/rde.2024.49.e21.

[50] Manganaro, L. The true latency of biomedical research papers. *Scientometrics* **2024**, *129*, 2897-2910, doi:10.1007/s11192-024-05008-0.

[51] Luwel, M.; van Eck, N.J.; van Leeuwen, T. Characteristics of Publication Delays Over the Period 2000–2016. In *Evaluative Informetrics: The Art of Metrics-Based Research Assessment : Festschrift in Honour of Henk F. Moed*, Daraio, C., Glänzel, W., Eds.; Springer International Publishing: Cham, 2020; pp. 89-114, doi:10.1007/978-3-030-47665-6_4.

[52] Gilliland, S.W.; Cortina, J.M. Reviewer and editor decision making in the journal review process. *Personnel Psychology* **2006**, *50*, 427-452, doi:10.1111/j.1744-6570.1997.tb00914.x.

[53] Lee, P.; Bubeck, S.; Petro, J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *N Engl J Med* **2023**, *388*, 1233-1239, doi:10.1056/NEJMsr2214184.

[54] Sallam, M.; Al-Salahat, K. Below average ChatGPT performance in medical microbiology exam compared to university students. *Frontiers in Education* **2023**, *8*, 1333415, doi:10.3389/feduc.2023.1333415.

[55] Sallam, M.; Khalil, R.; Sallam, M. Benchmarking Generative AI: A Call for Establishing a Comprehensive Framework and a Generative AIQ Test. *Mesopotamian Journal of Artificial Intelligence in Healthcare* **2024**, *2024*, 69-75, doi:10.58496/MJAIH/2024/010.

[56] Besancenot, D.; Huynh, K.V.; Faria, J.R. Search and research: the influence of editorial boards on journals' quality. *Theory and Decision* **2012**, *73*, 687-702, doi:10.1007/s11238-012-9314-7.

[57] Yeo-Teh, N.S.L.; Tang, B.L. An alarming retraction rate for scientific publications on Coronavirus Disease 2019 (COVID-19). *Account Res* **2021**, *28*, 47-53, doi:10.1080/08989621.2020.1782203.
[58] Huisman, J.; Smits, J. Duration and quality of the peer review process: the author's perspective. *Scientometrics* **2017**, *113*, 633-650, doi:10.1007/s11192-017-2310-5.
[59] Mikalef, P.; Gupta, M. Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management* **2021**, *58*, 103434, doi:10.1016/j.im.2021.103434.
[60] Sallam, M.; Al-Mahzoum, K.; Almutawaa, R.A.; Alhashash, J.A.; Dashti, R.A.; AlSafy, D.R.; Almutairi, R.A.; Barakat, M. The performance of OpenAI ChatGPT-4 and Google Gemini in virology multiple-choice questions: a comparative analysis of English and Arabic responses. *BMC Res Notes* **2024**, *17*, 247, doi:10.1186/s13104-024-06920-7.
[61] Herrmann-Werner, A.; Festl-Wietek, T.; Holderried, F.; Herschbach, L.; Griewatz, J.; Masters, K.; Zipfel, S.; Mahling, M. Assessing ChatGPT's Mastery of Bloom's Taxonomy Using Psychosomatic Medicine Exam Questions: Mixed-Methods Study. *J Med Internet Res* **2024**, *26*, e52113, doi:10.2196/52113.
[62] Raisch, S.; Krakowski, S. Artificial Intelligence and Management: The Automation–Augmentation Paradox. *Academy of Management Review* **2021**, *46*, 192-210, doi:10.5465/amr.2018.0072.
[63] Rampelotto, P.H. A Critical Assessment of the Peer Review Process in Life: From Submission to Final Decision. *Life* **2023**, *13*, 1603, doi:10.3390/life13071603.
[64] MDPI. Overview. Available online: https://www.mdpi.com/about (accessed on 4 September 2024).
[65] Oviedo-García, M.Á. Journal citation reports and the definition of a predatory journal: The case of the Multidisciplinary Digital Publishing Institute (MDPI). *Research Evaluation* **2021**, *30*, 405-419a, doi:10.1093/reseval/rvab020.

Supplementary Table

Comparison of human editor reasons for rejection as opposed to those provided by ChatGPT-4o ($n = 9$).

Ms#	Human editor	ChatGPT-4o
004	Your manuscript has been reviewed by the editors, who have decided to refuse the manuscript without sending it out for peer review. **XYZ** receives many high-quality submissions. Some of these submissions are not sent out for external review even though they may contain interesting or scientifically valid information. Our primary reasons for not soliciting external review may include the following: 1) The manuscript is better suited to a different journal. 2) The research findings or viewpoints are not sufficiently novel to warrant publication. 3) The paper would require revisions too extensive to be readily accomplished by the authors. Thank you for submitting your work for consideration.	While the manuscript touches on an interesting and relevant AI model performance for infectious disease queries—it does core focus of **XYZ**. The journal prioritizes clinical and trans studies that directly influence clinical practice in infectious dise addresses an important topic, its primary focus is on the perf different languages, which falls outside the journal's primary sc Furthermore, the study does not present a novel or significan infectious diseases. The findings are descriptive, highlighting disparities in AI, but do not offer actionable insights or in directly impact infectious disease management or patient care. Ethically, there are no significant concerns with the study its **XYZ**'s readership is limited. The manuscript might be better s in digital health, health informatics, or AI in medicine, wher audience more interested in AI model performance and its impl In conclusion, the manuscript does not meet the criteria for pub rejected. The authors are encouraged to consider submitting t in healthcare or digital health, where the content would be mor
005	Your manuscript entitled "**xyzxyz**" has now been assessed. If there are any reviewer comments on your manuscript, please find them below. Regrettably, the above submission has been rejected for publication in **XYZ**. We have made the decision to reject your paper without review since that ensures you received a decision now. This can give you an opportunity to submit the work at a more suitable venue. The novel contribution of this work is out of scope for **XYZ**, which currently focuses on the research into the application of artificial intelligence techniques and concepts to the design and evaluation of systems to support human learning and education. Thank you for the opportunity to consider your work. I am sorry that we cannot be more positive on this occasion and hope you will not be deterred from submitting future work to **XYZ**.	While the manuscript presents an interesting study on the fact attitudes towards ChatGPT among university students in the with the primary aims and scope of the **XYZ**. The journa application of AI to education, with an emphasis on the d designing computer-based learning systems, including topic environments, intelligent tutoring systems, and cognitive mode This manuscript, however, centers on the acceptance and (ChatGPT) rather than on the development, evaluation, or th systems designed for educational purposes. The study is m technology acceptance and usage patterns, which is better su educational technology adoption, user behavior, or social impact Additionally, while the manuscript appears to follow meth contribute novel insights into the AI-driven educational syste central to **XYZ** 's focus. There are also no significant theoretica the design of AI-based educational tools or systems. Given these considerations, I recommend rejection of the subm the core topics that **XYZ** seeks to explore. The authors may co that focuses on technology acceptance or educational technolog
006	Thank you for giving us the opportunity to consider your work. Unfortunately, your work doesn't fit in the scope of the journal. I'd like to suggest that you submit your work to another journal.	While the manuscript presents an interesting bibliometric a ChatGPT publications, it does not align well with the aims an focuses on metadata applications, interoperability, controlled retrieval within libraries and related institutions.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

		Although bibliometric analysis is a valuable research tool, the healthcare-related publications and the impact of ChatGPT fal matter of metadata in library science. Additionally, the ma contribute new knowledge or practices in the area of meta standards, or digital libraries. It centers more on the anal alternative metrics within a specific research domain rather application of metadata schemas or frameworks, which are cent The ethical considerations and methodological rigor of the however, due to the misalignment with the journal's scope, it submission. The authors may consider submitting this manus on bibliometrics or healthcare informatics, where the study's findings would be more relevant and impactful.
007	Thank you for submitting the following manuscript to **XYZ** We are writing to inform you that we will not be able to process your submission further. Submissions sent for peer-review are selected on the basis of discipline, novelty and general significance, in addition to the usual criteria for publication in scholarly journals. Therefore, our decision is not necessarily a reflection of the quality of your work. We wish you every success if you choose to submit it elsewhere.	The manuscript's focus on evaluating the output of ChatGPT conspiracies and compulsory vaccination presents an interesti given the increasing reliance on AI for information retrieva several key areas: 1. **Relevance to Journal Scope**: While the manuscript t vaccine-related topics, its primary focus is on the performanc advancing vaccine research, immunology, or vaccine technol scope of " **XYZ**." The study's emphasis on AI evaluation o immunological mechanisms makes it less relevant to the core au 2. **Methodological Rigor**: The study lacks a robust methodo single AI session (conducted on a specific date) and evaluation provide sufficient depth or reproducibility. The assessment of qualitative analysis with Cohen's kappa is too narrow and la study aiming to contribute to scientific discourse. 3. **Ethical Considerations**: The study does not address impo to AI in healthcare, such as potential misinformation, the li medical advice, or the need for human oversight. These eth evaluating AI tools in a public health context. 4. **Novelty**: The manuscript does not provide significant no advance the field of vaccine research or public health. Instead, issues with AI and online misinformation without offering new Given these considerations, the manuscript does not align suffic of " **XYZ** " and lacks the necessary scientific and ethical rigor to be sent for peer
008	Thank you for submitting your manuscript for consideration for publication in **XYZ**. An expert in the field and the editor have reviewed your manuscript and found it unsuitable for publication. The main criticisms are the lack of focus, the research design, the method used, and the study's presentation. Medical education is missing from the manuscript. I have enclosed the Reviewers' Reports below with my letter. This decision applies to all BMC journals. This decision should not stop you from submitting your future research to **XYZ**.	While the study is methodologically sound and presents a validating a survey instrument based on the Technology Acce not align well with the specific scope and aims of * **XYZ** *. university educators' perspectives on ChatGPT, a tool with l medical field. Although the study offers valuable insights technologies in higher education, it lacks a direct and clear imp which is a core requirement for publication in ***XYZ***. The study also has limited generalizability due to its regional f could be acceptable if the findings had broader implications, th make a sufficiently strong case for the relevance of its finding education community. Additionally, while AI- related studies this study focuses more on the validation of an instrument for a rather than on the impact or application of AI in medical educat

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

	Thank you for the opportunity to review your work. I'm sorry that we cannot be more positive on this occasion and hope you will not be deterred from submitting future work to BMC Medical Education.	from the journal's main interests. In conclusion, although the study is scientifically valid and well for a journal that focuses on educational technology or higher is specifically centered on medical education. Therefore, the ma the grounds of relevance to the journal's scope.
	Reviewer Comments: Reviewer 1 Thank you for submitting the revised version of your paper. The quality of this paper has increased significantly after addressing the reviewers' comments. No further comments are required.	
009	Regrettably, your manuscript has been rejected for publication in **XYZ**. Thank you for the opportunity to review your work. I'm sorry that we cannot be more positive on this occasion and hope you will not be deterred from submitting future work to **XYZ**. Reviewer 1 Interesting themw The manuscript seems interesting, however number of records are too less to draw any significant conclusion. However, Author have analysed the publications with some good insights and statistics. It would be interesting for researchers and especially students go through this publication . I recommend it to be published in this journal Reviewer 2 Although the work put into the publication is much appreciated, there is still room for improvement in terms of outlining the consequences of the findings, even after a thorough reading. 1. An effort is made to identify many crucial bibliometric analysis criteria. 2. The literature review may go farther to comprehend the research issue. 3. The way the findings are presented may be made better. 4. The conclusion may be revised to more fully tie everything together. Recommendation: Major Revision	While the manuscript addresses a timely and relevant topic, th well with the transdisciplinary nature and forward-looking m centers on a bibliometric analysis of ChatGPT publications in h does not offer new theoretical insights, methodologies, or sign journal seeks to publish research that proposes innovative ap transformative impact of AI in various domains. This stu retrospective analysis of publication metrics, which may be bett on bibliometrics or the sociology of science. From a methodological standpoint, the paper appears to be rigo multiple databases and metrics to assess citation impact. How the study simply ranks publications based on citation counts an metrics. The findings, while interesting, do not advance the u AI in healthcare. Ethically, the study acknowledges limitations but does not see might influence or contribute to the ethical use of AI in health call for collaborative initiatives to establish guidelines, while in way that offers new ethical frameworks or solutions, which is a Given these considerations, the manuscript does not meet the and contribution to the field required for publication in * **XYZ** *.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

010	Your manuscript entitled " **xyzxyz** " has now been reviewed. Any reviewer comments on the suitability of your manuscript have been appended below. As a result, I regret to inform you that we cannot publish your manuscript in **XYZ**. You will see that, while your work is of interest, substantive concerns were raised that suggest that your paper does not fulfil the publication requirements for **XYZ** that is, that papers must be technically sound in	The manuscript addresses an important and timely topic—th models in the context of clinical chemistry education. However decision to reject this submission: 1. **Relevance to Journal Scope:** While the topic interse engineering, the primary focus appears to be on the performa rather than advancing knowledge in clinical chemistry or a Scientific Reports aims to publish original research that con natural sciences, psychology, medicine, and engineering, but towards a comparative study of AI models, which might b specializing in AI or education.
	method and analysis. Unfortunately, these reservations are sufficiently important to preclude publication of this study in **XYZ**. Editor comments The article provides preliminary data and insights, but does not fully demonstrate how these analyses can be translated into clinical applications, and there is insufficient discussion on technological maturity. In addition, the small sample size used may limit the statistical power and reliability of the research results. Thank you for the opportunity to consider your work. I am sorry that we cannot be more positive on this occasion and hope you will not be deterred from submitting future work to **XYZ**.	2. **Novelty and Impact:** The manuscript compares several does not provide novel insights into the mechanisms or implica The focus on MCQs as an evaluation tool is not sufficiently i the broader scope of healthcare and education research. Th interesting, do not advance understanding in a way that ju impact, multidisciplinary journal. 3. **Methodological Rigor:** While the study adheres to the the CLEAR tool for quality assessment, the reliance on a limi concerns about the robustness and generalizability of the findi from a more extensive dataset and a broader evaluation fr aspects of AI model performance, such as reasoning or clinical c 4. **Ethical Considerations:** The manuscript raises vali dishonesty and dependence on AI, but these are not explore ethical implications of using AI models in education, particul the development of critical thinking skills in students, should b Overall, while the manuscript is well-structured and addresses in terms of novelty, impact, and alignment with the journ exploration of the implications of AI in healthcare education, methodological rigor and ethical considerations, would be neces
014	Your manuscript entitled " **xyzxyz** " has now been assessed. If there are any reviewer comments on your manuscript, please find them below. Regrettably, the above submission has been rejected for publication in **XYZ**. We have made the decision to reject your paper without review since that ensures you received a decision now. This can give you an opportunity to submit the work at a more suitable venue. This work concerns a very timely and important topic, the ways to make effective use of emerging generative AI technology for education. Indeed, there are many valuable blogs and posts on the web on this topic. However, the current paper does not reflect sufficient maturity for **XYZ**. One key difficulty	While the manuscript presents an intriguing study comparing chatbots and human students on Clinical Chemistry multiple does not align well with the specific aims and scope of the * **XY** the application of AI in education, particularly on the developm computer-based learning systems. The study, however, prima of AI models on a specific set of MCQs without sufficient implications or the design principles of AI-based learning syste **Scientific Rationale:** - **Relevance:** The manuscript focuses on the performance human students in answering MCQs. However, it does not informs the design or improvement of AI- driven educational to - **Novelty:** While the topic of AI outperformance is relevant novel concepts or methodologies in the realm of AI in edu aligned with AI performance benchmarking rather than a education. - **Clarity and Methodological Rigor:** The study is appropriate guidelines (e.g., METRICS checklist), but the practices and the development of intelligent tutoring systems a

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189

	is that the particular versions of the tools, such as ChatGPT, have a substantial impact on the results of experiments such as those you report. Without the deep analysis of the nature of the tools and the implications of this, the contribution of this work is limited. Thank you for the opportunity to consider your work. I am sorry that we cannot be more positive on this occasion and hope you will not be deterred from submitting future work to **XYZ**.	**Ethical Considerations:** - **Academic Dishonesty:** The manuscript touches on the dishonesty with AI use, but it does not propose actionable solut guide educators in mitigating these risks within educational s relevant to the journal's audience. In conclusion, the manuscript does not sufficiently address the way that aligns with the core mission of **XYZ**. Therefore, it is recommended to
029	Thank you for submitting the following manuscript to **XYZ**	1. **Relevance and Scope:** The manuscript primarily focuses o D614G mutation and its phylogenetic analysis within SARS-CoV-2 sequences fr this topic is certainly
	We are writing to inform you that we will not be able to process your submission further. Submissions sent for peer-review are selected on the basis of discipline, novelty and general significance, in addition to the usual criteria for publication in scholarly journals. Therefore, our decision is not necessarily a reflection of the quality of your work. We wish you every success if you choose to submit it elsewhere.	relevant in the broader context of microbial genomics and epid focus on a single mutation (D614G) and phylogenetic analysis or novel insights into microbial physiology, genetics, or ecolog broader readership of * **XYZ** *. The journal explicitly exclu report single-gene phylogenies or descriptions of single spe largely does. 2. **Novelty and Contribution:** The D614G mutation has various global contexts, and the findings reported in this r provide significant new insights or advancements in our un evolution or its impact. The increase in the D614G mutation well-documented in multiple studies, limiting the novelty of this 3. **Methodological Rigor:** While the methodology (use of M phylogenetic analysis) appears sound, the study lacks depth in for viral evolution, epidemiology, or public health. The man detailed experimental or theoretical advancements that woul within this journal. 4. **Ethical Considerations:** No ethical issues are apparent, present a strong case for its necessity or urgency within the context of the curren Given these considerations, the manuscript does not meet the s publication in * **XYZ** *.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 960-979, 2024
DOI: 10.55214/25768484.v8i6.2189