

## Artificial intelligence to support publishing and peer review: A summary and review

Item Type	Journal article
Authors	Kousha, Kayvan;Thelwall, Mike
Citation	Kousha, K. and Thelwall, M. (2024) Artificial intelligence to support publishing and peer review: A summary and review. <i>Learned Publishing</i> , 37 (1), pp. 4-12. <a href="https://doi.org/10.1002/leap.1570">https://doi.org/10.1002/leap.1570</a>
DOI	<a href="https://doi.org/10.1002/leap.1570">10.1002/leap.1570</a>
Publisher	Wiley
Journal	Learned Publishing
Download date	2026-03-09 08:28:47
License	<a href="http://creativecommons.org/licenses/by/4.0/">http://creativecommons.org/licenses/by/4.0/</a>
Link to Item	<a href="http://hdl.handle.net/2436/625277">http://hdl.handle.net/2436/625277</a>



# Artificial intelligence to support publishing and peer review: A summary and review

Kayvan Kousha <sup>1</sup>, and Mike Thelwall <sup>1,2\*</sup>

<sup>1</sup>Statistical Cybermetrics and Research Evaluation Group, University of Wolverhampton, Wolverhampton, UK

<sup>2</sup>Information School, University of Sheffield, Sheffield, UK

ORCID:

K. Kousha: [0000-0003-4827-971X](https://orcid.org/0000-0003-4827-971X)

M. Thelwall: [0000-0001-6065-205X](https://orcid.org/0000-0001-6065-205X)

\*Corresponding author: Mike Thelwall, Information School, University of Sheffield, UK.  
E-mail: [m.a.thelwall@sheffield.ac.uk](mailto:m.a.thelwall@sheffield.ac.uk)

**Abstract:** Technology is being developed to support the peer review processes of journals, conferences, funders, universities, and national research evaluations. This literature and software summary discusses the partial or complete automation of several publishing-related tasks: suggesting appropriate journals for an article, providing quality control for submitted papers, finding suitable reviewers for submitted papers or grant proposals, reviewing, and review evaluation. It also discusses attempts to estimate article quality from peer review text and scores as well as from post-publication scores but not from bibliometric data. The literature and existing examples of working technology show that automation is useful for helping to find reviewers and there is good evidence that it can sometimes help with initial quality control of submitted manuscripts. Much other software supporting publishing and editorial work exists and is being used, but without published academic evaluations of its efficacy. The value of artificial intelligence (AI) to support reviewing has not been clearly demonstrated yet, however. Finally, whilst peer review text and scores can theoretically have value for post-publication research assessment, it is not yet widely enough available to be a practical evidence source for systematic automation.

**Keywords:** academic peer review, academic publishing, artificial intelligence, automation

## INTRODUCTION

Within the academic publishing system, complex decisions are made tens of millions of times annually by highly trained researchers, raising the possibility that substantial efficiency improvements could be made by software that would automate or otherwise support them. The decision makers include authors (Is this manuscript written and formatted appropriately? Which journals are suitable publication venues for it?); editors (Is this submission plagiarised? Is it in scope? Does it have clear style/

guideline/methodological flaws? Who are suitable reviewers for it?); reviewers (Does the submission contain clear errors? What is its quality? What changes should be recommended?); editors again (What decision should be made from the reviewer reports?); and research managers (How good is the published paper?). This review discusses existing software to support these tasks.

The time savings from automation are potentially huge because reviewing alone consumes a large amount of expert time. It was estimated that 'over 15 million hours' are spent on reviewing *rejected* papers each year (American

Journal Experts, 2018). For example, 1.2 million manuscripts are submitted to 2300 Elsevier journals every year and only 30% (about 350,000) are published (Tedford, 2015). Due to increasing numbers of submissions and peer review workloads, a report from BioMed Central and Digital Science entitled ‘What might peer review look like in 2030?’, recommended to ‘use technology to support and enhance the peer review process, including finding automated ways to identify inconsistencies that are difficult for reviewers to spot’ (Burley & Moylan, 2017, p. 3). For instance, about 20% of biomedical researchers conducted ‘69% to 94% of the reviews’ in 2015 (Kovanis et al., 2016).

This review discusses artificial intelligence (AI) and other software to fully or partially automate aspects of the academic publication process, including post-publication quality evaluation. It starts with software to recommend journals to authors, then covers programs for initial quality control of submitted manuscripts (e.g., plagiarism detection), followed by initiatives to suggest appropriate reviewers for papers that have not been desk rejected and then programs to conduct reviews or make recommendations from human reviewers’ reports or scores. Finally, the review discusses software to estimate the quality of submitted articles or predict their future citation counts from peer review text or scores. Research quality assessment and citation impact prediction from bibliometric data is not covered here. Bias and transparency in technology-assisted assessment are also not in scope (all are covered in separate reviews), although software to detect bias and transparency within manuscripts is mentioned. This paper primarily summarizes existing software, including where there is no academic evidence of its capabilities, but also analyses relevant academic research in the minority of cases where it exists—primarily for reviewer identification and post-publication peer review.

## IDENTIFYING SUITABLE JOURNALS FOR A SUBMISSION

Identifying a relevant journal or conference is a significant step towards publishing research because a good article might be rejected from an inappropriate journal or find a smaller audience. Several publishers have developed web services to suggest journals for manuscripts, either for initial submissions or to redirect rejected manuscripts to another of the publisher’s journals. This software seems to be mainly based on comparing article texts (titles, abstracts, or keywords) against previously published articles. Current journal recommendation tools include *Springer Nature Journal Suggester*,<sup>1</sup> *Wiley Journal Finder*<sup>2</sup> or *IEEE Publication Recommender*.<sup>3</sup> *EndNote Manuscript Matcher* also uses a manuscript’s title, abstract, and references with Web of Science data to suggest related journals for manuscripts.<sup>4</sup> *The Journal/Author*

<sup>1</sup><https://journalsuggester.springer.com/>

<sup>2</sup><https://journalfinder.wiley.com/search?type=match>

<sup>3</sup><https://publication-recommender.ieee.org/home>

<sup>4</sup><https://endnote.com/product-details/manuscript-matcher/>

### Key points

- Artificial Intelligence in peer review is useful for helping to find reviewers.
- Artificial Intelligence can sometimes help with initial quality control for manuscripts submitted to academic journals.
- This article reviews the use of Artificial Intelligence to support academic peer review and publishing.
- Artificial Intelligence does not seem yet capable of performing peer review for manuscripts submitted to academic journals.

*Name Estimator (JANE)* is another free service that uses text similarity scores and PubMed data to suggest the most relevant journals based on manuscript titles or abstracts.<sup>5</sup> Similarly, *JournalGuide*<sup>6</sup> from Research Square claims to recommend a journal from a title and abstract.

Various types of AI are used in these systems to match the subjects of manuscripts to related journals. For instance, *Elsevier’s JournalFinder*<sup>7</sup> service ‘uses smart search technology and field-of-research specific vocabularies’ to match paper to scientific journals; the *Taylor & Francis’ Journal Suggester*<sup>8</sup> applies ‘artificial intelligence to match the subjects covered in articles’; and the *Sage Journal Selector*<sup>9</sup> utilizes ‘an advanced AI technology’ to recommend journals with similar published articles.

Several studies have shown that AI or machine learning can be useful to identify appropriate academic journals or conferences with relatively high accuracy for papers (e.g., Feng et al., 2019; Ghosal, Raj, et al., 2019; Pradhan & Pal, 2020; Wang et al., 2018). For instance, a recent experiment used the XGBoost algorithm and different features (title, abstract, and keywords) from 20,250 articles from Web of Science-indexed computer technology journals, reporting an accuracy of 84% for academic journal recommendations (ZhengWei et al., 2022). Using deep learning techniques, another study developed a journal recommendation system (*Pubmender*) to propose appropriate PubMed journals based upon articles’ abstracts. The experiential dataset included abstracts of over 880,000 papers from 1130 PubMed Central journals, reporting an accuracy of 87%. This was claimed to be much higher than Elsevier’s *Journal Finder* and Springer’s *Journal Suggester* tools (Feng et al., 2019). *GraphConfRec* has also been developed to recommend relevant computer science conferences based on paper text, co-authorship, and citation networks (Iana & Paulheim, 2021).

<sup>5</sup><https://jane.biosemantics.org/>

<sup>6</sup><https://www.journalguide.com/>

<sup>7</sup><https://journalfinder.elsevier.com/>

<sup>8</sup><https://authorservices.taylorandfrancis.com/publishing-your-research/choosing-a-journal/journal-suggester/>

<sup>9</sup><https://www.edanz.com/journal-selector>

## INITIAL QUALITY CONTROL FOR SUBMITTED MANUSCRIPTS

AI and Natural Language Processing can assist editors and publishers in many ways, from sending automatic emails to authors or reviewers and checking plagiarism in the manuscripts to statistical test and methods checking (for reviews see: Checco et al., 2021; Lin et al., 2021; Price & Flach, 2017). One study compared different manuscript management tools (e.g., *ScholarOne*, *Editorial Manager*, *EVISE* and *Open Journal Systems*) from the authors', reviewers', and editors' perspectives, reporting tools supporting automatic editorial tasks (e.g., sending e-mails, reviewer recommendation, statistical analysis, similarity check, linking references to Crossref or/and PubMed; Kim et al., 2018). Tools that can automate some tasks of editorial management or the peer review process include the following.

- **Plagiarism detection:** *iThenticate*<sup>10</sup> detects partially or fully copied text in a manuscript by comparing it to other manuscripts in the system (Kalnins et al., 2015). A procedure to detect tortured phrases can identify plagiarism in the form of papers or sections of papers that have been processed by paraphrasing software or that have been translated from English to another language and back again. It identifies meaningless phrases that are non-idiomatic translations of scientific phrases, such as 'counterfeit consciousness' from 'artificial Intelligence' (Cabanac et al., 2021).
- **Robot author detection:** A more recent threat is the use of AI based on large language models (LLMs) like ChatGPT to write parts of articles in a way that may be plausible but lacks deep understanding of the research issues (Khalil & Er, 2023). Software to detect ChatGPT-authored text, such as ZeroGPT,<sup>11</sup> although not yet designed for manuscript reviewing, may help with this issue.
- **Methods checking:** *SciScore*<sup>12</sup> generates an automated assessment of articles methods on a scale of 1–10 and other reports (Design Analysis Reporting checklist and the Rigor and Transparency Index), assisting reviewers to find key information throughout a paper in a standard format (see also Menke et al., 2020). *RobotReviewer*<sup>13</sup> attempts to detect study design problems and bias in randomized controlled trials.
- **Automated statistical checking:** *StatReviewer*<sup>14</sup> checks manuscripts against standardized reporting guidelines and *StatCheck*<sup>15</sup> detects some statistical errors in the submitted works (Nuijten & Polanin, 2020). It is possible to check the plausibility of some statistical results in papers by testing if the numbers reported for a test are theoretically capable of having been generated at the level of rounding reported.

<sup>10</sup><http://www.ithenticate.com>

<sup>11</sup><https://www.zerogpt.com/>

<sup>12</sup><https://www.sciscore.com/>

<sup>13</sup><https://www.robotreviewer.net/about>

<sup>14</sup><http://www.statreviewer.com/>

<sup>15</sup><http://statcheck.io/>

- **Transparency and reproducibility checking:** *Dimensions Research Integrity preCheck*<sup>16</sup> analyses submitted manuscripts for evidence of transparency and reproducibility, such as data access statements and naming the versions of software used.
- **Manuscript structure checking:** *Penelope.ai*<sup>17</sup> checks if the structure of a manuscript meets a journal's submission guidelines for the title page, abstract, citation style, references, tables and figures and information about other sections of articles (e.g., funding, acknowledgements, keywords, and data/ethics statements). This reduces the need for manual checks by reviewers, publishers, or editors.
- **Reference matching with in-text citations:** *Recite*<sup>18</sup> automatically checks and highlights if citations in the manuscript text match the reference list and vice versa
- **Multipurpose manuscript evaluation:** Other AI-assisted tools assess multiple quality control aspects of manuscripts such as *Frontiers AI Review Assistant*<sup>19</sup> or *UNSILO Manuscript Evaluation*<sup>20</sup> (see also Heaven, 2018). *AuthorOne*<sup>21</sup> performs multiple surface level checks on manuscripts to see if they are written and formatted appropriately.

## IDENTIFYING APPROPRIATE REVIEWERS FOR A SUBMISSION

Selecting an appropriate set of reviewers for a submitted journal article, conference paper, or grant proposal is an important and challenging task. It requires knowledge of the skills needed to assess the work as well as the people that possess those skills. The same is true for large-scale post-publication peer review exercises such as the UK's Research Excellence Framework (REF), because the 34 subpanel chairs (senior professors) must start the review process by assigning each output (185,594 for REF2021; REF, 2022) to at least two reviewers, an enormous task. Fully or partly automating this labour-intensive process might improve the overall match between subpanel reviewers and articles and save time. Editorial systems for publishers and some conferences already support this task by suggesting possible reviewers for submitted articles, perhaps based on references in the submitted outputs or by matching article keywords to the keywords of registered reviewers. Conference-based reviewer bidding systems are also sometimes used to allow reviewers to choose articles that they would like to review, with algorithms or human judgements selecting the final set of reviewers (Fiez et al., 2020).

Several programs have been developed by commercial publishers to help journal editors identify suitable reviewers, but

<sup>16</sup><https://ripeta.com/>

<sup>17</sup><https://www.penelope.ai/faq>

<sup>18</sup><https://reciteworks.com/>

<sup>19</sup><https://blog.frontiersin.org/2020/07/01/artificial-intelligence-peer-review-assistant-aira/>

<sup>20</sup><https://unsilo.ai/unsilo-manuscript-evaluation/>

<sup>21</sup><https://www.authorone.ai/>

these do not seem to have publicly available algorithm descriptions or evaluations.

- *Clarivate's Reviewer Locator*<sup>22</sup> automatically suggests reviewers based on data from the Web of Science and Publons peer review databases and connects to the ScholarOne submission management system integrating editorial and peer-review processes.
- *Reviewer Discovery*<sup>23</sup> is a tool from Aries Systems that uses ProQuest author profiles and automatically suggests reviewers based on the Editorial Manager system.
- *Elsevier's EVISE*<sup>24</sup> uses Reviewer Finder to identify and recommend reviewers based on Scopus data.
- *Frontiers Coronavirus Reviewer Recommender*<sup>25</sup> suggests experts to review COVID-19 research proposals using 'Frontiers knowledge graph and advanced information extraction and retrieval methods'.

The same principles that work to identify journal reviewers can also work for research grants. The Natural Science Foundation of China (NSFC) has developed an AI-assisted reviewer recommender for grant applications using natural language processing and an assignment decision support system to help select expert panels. An initial version of the AI system had chosen 'at least one member of each of nearly 44,000 panels that approved projects' in 2018, and the accuracy of the system was about 80% (Cyranoski, 2019, p. 317), although accuracy improvements are still being made (Liu et al., 2022). The system classifies the reviewers and proposals by discipline and uses information from scientific databases (e.g., Web of Science) and referee profiles in NSFC databases about the publication records or research projects of potential reviewers and then uses lexical semantic analysis to compare the extracted information with the grant applications. Different rules were used in the system to avoid conflicts of interests between reviewers and applicants (e.g., affiliation, co-authorship, project, and tutor-student relationships; Liu et al., 2016).

Automatic reviewer assignment is also possible for conferences. The Toronto Paper Matching System automatically suggested reviewer assignments for the NIPS 2010 conference using a topic modelling approach to estimate reviewers' expertise areas. The system extracts publication records from Google Scholar to generate profiles for reviewers and uses supervised score prediction models to suggest reviewer assignments (Charlin & Zemel, 2013). Several other studies have also suggested algorithms for the automatic assignment of reviewers to conference papers (e.g., Al Mahmud et al., 2018; Kalmukov, 2020; Li & Watanabe, 2013), mostly for AI-related

conferences. No robust accuracy measures seem to have been generated for these systems yet, however. Presumably the ground truth for such a system would be human editor assignments or (in conferences that allow this) reviewer requests to review. A simulation platform has also been developed to generate artificial review threads, reducing peer review time by about 30% (Mrowinski et al., 2017).

Automation does not seem to have been used yet to find reviewers in post-publication peer review exercises like the REF, but the technologies described above could presumably achieve this quite easily, at least in journal-based fields.

## REVIEWING A SUBMISSION

Conducting peer review can be a fruitful learning experience for the reviewer but is usually primarily considered by reviewers to be a necessary task to contribute to the scientific community (Zaharie & Osoian, 2016), in addition to its quality control, improvement, and educational roles (Resnik & Elmore, 2016). As such, any technological solutions to save time in peer review or associated editorial work could create an immediate tangible benefit to science.

Although some software has been developed to review papers or make editorial decisions, these tasks are challenging, with limited progress on a few topics so far. Whilst positive correlations between human and automated decisions have been generated, no current system challenges human reviewing yet. The positive correlations between peer review judgements and machine learning found so far do not necessarily mean that further progress is likely soon because an AI system would achieve a positive correlation by rejecting papers with obvious errors, such as very poor grammar, too short, or lacking references.

*ReviewAdvisor*<sup>26</sup> is a natural language processing toolkit designed to help select good manuscripts for a journal and provide feedback to help authors improve their submitted articles. Whilst its performance on the authors' ASAP-review set of 28,119 machine learning conference paper reviews was weak, it provides a starting point and might help reviewers by suggesting comments on paper aspects that they may have overlooked. Another study developed a neural network AI tool trained on features from papers including word frequencies, readability scores, and formatting measures, finding that automated systems developed unethical biases, such as against grammar and formatting errors, that helped them be more accurate (Checco et al., 2021). Similarly, the *pReview* software package was developed for automatically generating summarization, contribution detection, writing quality analysis, and potential related works of academic papers to support reviewers (Roberts & Fisher, 2020). Another study found that Natural Language Processing models to generate reviews for scientific papers could make the peer review task easier and more effective, but the reviews were not good enough to replace human experts (Yuan et al., 2021).

<sup>22</sup><https://clarivate.com/webofsciencegroup/solutions/reviewerlocator/>

<sup>23</sup><https://www.ariessys.com/newsletter/june-2013/reviewer-discovery-free-trial-available/>

<sup>24</sup><https://www.elsevier.com/connect/reviewers-update/rolling-out-our-new-editorial-system-evise>

<sup>25</sup><https://coronavirus.frontiersin.org/covid-19-research-funding-monitor>

<sup>26</sup><https://github.com/neulab/ReviewAdvisor>

Review decision-making software makes recommendations without writing detailed comments. For example, a study used 3,341 papers from three computing conferences (the 2018 IEEE Wireless Communications and Networking Conference and the 2018 and 2019 International Conference on Learning Representations) and assessed to what extent AI can predict human peer review decisions about papers (acceptance/rejection or average review score). Using a Random Forest classifier, the above study predicted the acceptance of papers submitted to a top AI conference with 81% accuracy based on surface features, such as the number of tables in the papers or characters in the title (Skorikov & Momen, 2020).

General purpose LLMs like ChatGPT have potential reviewing applications, such as suggesting grammar corrections or perhaps even strategies to help authors correct problems identified by reviewers (Lund et al., 2023). Nevertheless, reviewers normally agree to keep the manuscripts assessed confidential; as such, papers under assessment should not be uploaded to LLMs because they may be saved and incorporated into responses to future questions for other users (Hosseini & Horbach, 2023), or may even grant the LLM owner the right to repurpose the content. Similarly, some authors may wish to avoid using LLMs to protect their ideas.

## MAKING REVIEW DECISIONS FROM PEER REVIEW COMMENTS

After reviewers have submitted their recommendations and comments, the next stage is for the journal editor, grant award board, or conference organizers to make a final decision. Since it is common for reviewers to disagree and their recommendations might not align with their comments due to inadequate norm referencing, systems to automate final judgements might be helpful, especially if they can give reasons for their decisions that can be checked.

Based on a dataset of scientific peer reviews from *PeerRead*,<sup>27</sup> a deep learning network has been used to predict the acceptance or rejection of articles from peer review reports and to generate the final meta-review, finding 'good consistency between the recommended decisions and original decisions', with 74%–86% accuracy at predicting the binary decision accept/reject, which was better than standard machine learning algorithms and prior bespoke peer review judgement algorithms (Pradhan et al., 2021, p. 237). There is also evidence that sentiment analysis of review reports could be helpful to predict the final decision (acceptance or rejection) of conference papers (Chakraborty et al., 2020; Ghosal, Verma, et al., 2019; Wang & Wan, 2018) or review scores of funding programmes (Luo et al., 2021). For example, *PeerJudge*<sup>28</sup> uses AI-assisted sentiment detection to estimate the strength of praise and criticism in peer

review reports on academic papers that could be useful for editorial management decisions based on analysing a large number or review reports. *PeerJudge* can predict F1000Research reviewer decisions with a moderate degree of accuracy (Thelwall et al., 2020).

## POST-PUBLICATION PEER REVIEW ANALYSIS

After the formal publication of research outputs, an increasing minority have peer review reports or post-publication comments associated with them. This can occur because the publishing journal or conference either mandate (e.g., BMJ, BMC Cancer) or allow (e.g., MDPI journals) peer review reports to be published, because the reviewer published their reports elsewhere, or because other academics posted comments to the publishing website (e.g., BMJ, F1000Research) or a bespoke academic commenting website (e.g., *PubPeer*). There are some parallel initiatives to move to a publishing first model in which preprints are first posted and then reviews are solicited either through editorial requests (e.g., F1000Research) or on a voluntary basis. Pre- and post-publication reviews may help universities to access formal peer review reports to 'improve and promote research excellence assessment' (Wilkinson & Down, 2018). This also raises the possibility that in the future the public availability of peer review comments will increase to the extent that automated processing of them could be useful in post-publication research assessment, such as the UK REF. This section reviews related work in the hope that this will happen.

A technical problem with the automatic processing of peer review reports is that they are often difficult to parse for text mining because of their format, such as unstructured PDFs with extra text or commented copies of reviewed article PDFs or Word documents (e.g., Thelwall et al., 2022). A deeper problem is that most open reviews are from reviewers and address pre-final versions of the article, so it is not clear that they provide useful information about the final published article. Moreover, if public peer reviews are used in formal research assessments, then this might incentivize manipulation of them.

## Sources of open peer review

There are several open review platforms sharing pre-publication reviews (formal reviews or editorial comments from the publishing journal) or post-publication comments (recommendations or feedback by researchers or experts) for scholarly publications. *Publons* ([publons.com](http://publons.com)) was an open review platform, now taken over by Clarivate Analytics, that claimed to include 'over 6.9-million reviews for more than 5,000 partnered journals'.<sup>29</sup> Partner journals could share pre-publication reviews publicly on this site, with reviewers and authors deciding what information to reveal (e.g., review text, reviewer identities). In contrast, *PubPeer* ([pubpeer.com](http://pubpeer.com)) is an online open platform that focuses on post-publication

<sup>27</sup><https://paperswithcode.com/paper/a-dataset-of-peer-reviews-peerread-collection>

<sup>28</sup><http://sentistrength.wlv.ac.uk/PeerJudge.html>

<sup>29</sup><https://publons.freshdesk.com/support/solutions/articles/12000012231-what-is-publons-and-why-partner-with-us->

peer review, where researchers can provide feedback or comments about published research and authors can respond. Reviews in PubPeer have identified mistakes published in leading cell biology journals,<sup>30,31</sup> suggesting that post-publication reviews might be a helpful source for quality assessment of published research. A study of a sample of PubPeer comments about publications found that two-thirds were related to some type of misconduct (Ortega, 2022). *ScienceOpen* ([scienceopen.com](https://www.scienceopen.com)) combines publishing and promotion services for journals with a recommendation capability where other researchers or experts can write public reviews and use a five-star score about the 'importance', 'validity', 'comprehensibility', and 'completeness' of published research.<sup>32</sup> *Peer Community in* [peercommunityin.org](https://www.peercommunityin.org) is a free recommendation platform for preprints. It publishes peer reviews of preprints in 14 subject areas, including Ecology, Genomics, Animal Science, and Evolutionary Biology. The peer review process is managed by 1700 'Recommenders' making editorial decisions about public reviews.<sup>33</sup>

The Multidisciplinary Digital Publishing Institute (MDPI) and several other publishers and journals provide an open peer review option, where authors can decide to publish their reviews and reviewers can choose to be named or remain anonymous (for a review see Wolfram et al., 2020). These provide a collective source of open peer review reports for the journals covered. This can be used for systemic quality control purposes. For example, a study of 45,385 open standard article reviews for 288 MDPI journals found large disciplinary differences in review lengths, reviewer anonymity, review outcomes, and the use of attachments. For instance, reviewers in the Physical Sciences are most likely to ask for major revisions and to use attachments in the review process, although they are less likely to disclose their identity. In the Life Sciences and Social Sciences fields, reviewers tend to write longer review reports than in the Physical Sciences (Thelwall, 2022). The rest of this subsection reviews research into a former and a current major peer review site to illustrate how this type of content might be helpful for research assessment systems, including systemic self-assessment.

## Publons

[Publons.com](https://publons.com) was a website to help academic reviewers record and display their peer review work and editorial roles. It served to promote peer reviewing by tracking it publicly so that academics could actively or passively use it as evidence of their reviewing contribution to science, which would otherwise be invisible. In August 2022, the website started to redirect to the Web of Science and, at the time of writing, its basic free services seemed to be no longer available. Nevertheless, several studies

<sup>30</sup><https://phys.org/news/2013-05-stem-cell.html>

<sup>31</sup><https://www.nature.com/articles/nature.2013.13060>

<sup>32</sup><https://blog.scienceopen.com/2016/05/peer-review-at-scienceopen-is-surprisingly-simple/>

<sup>33</sup><https://peercommunityin.org/about/>

took advantage of the information it provided before August 2022 to investigate characteristics of peer review.

A study of 45,819 articles from Publons found low or insignificant correlations between bibliometric scores (e.g., WoS or Scopus citations) and Publons metrics (e.g., Quality, Significance and Overall Publons score of articles; Ortega, 2019; see also Ortega, 2017). For four small experimental groups of papers from Publons with neutral, negative, positive, and both negative and positive post-publication reviews, papers with positive reviews had significantly more citations ( $\rho = 0.498$ ,  $p < 0.05$ ) while very low or non-significant associations were found between citation counts and other review polarities (Zong et al., 2020). Thus, peer review metrics might not be useful indicators of citation impact or, by extension (because the two correlate in many fields), research quality. The availability of open peer review reports varies substantially between journals and fields (Ortega, 2019), and between reviewer countries (Severin et al., 2021), further undermining its value as an input for AI.

## Faculty opinions (formerly F1000Prime)

In contrast to the above, Faculty Opinions ([facultyopinions.com](https://www.facultyopinions.com), previously F1000 Prime) is a paywalled source of post-publication biomedical research reviews written 'by over 8,000 experts in the Life Sciences and Medicine'. Articles are classified based on contribution type, such as 'Good for Teaching', 'New Finding', 'Technical Advance,' or 'Interesting Hypothesis' and can be given one (Good), two (Very Good) or three (Exceptional) stars.<sup>34</sup> The article recommendations are exclusively positive: 'Good' (58.6%), 'Very Good' (34.6%), and 'Exceptional' (6.9%) (Waltman & Costas, 2014).

Several investigations have reported significant positive associations between Faculty Opinions ratings and citation metrics, suggesting that they have value as quality or impact indicators (Bornmann, 2015; Bornmann & Leydesdorff, 2013, 2015; Du et al., 2016; Li & Thelwall, 2012; Mohammadi & Thelwall, 2013; Smith et al., 2019; Waltman & Costas, 2014), with one early exception (Wardle, 2010). For example, for 830 research articles published in 2010 by four high-impact journals, those recommended in Faculty Opinions on average had more citations (between 2010 and 2019) and this difference was statistically significant for articles published in *Nature Genetics*, *Nature Medicine*, and *PLoS Biology* except for the journal *Cell* (Wang & Su, 2021).

Faculty Opinions ratings have also been shown to align with other expert opinions, giving more direct evidence of their value than the citation-based analyses above. A study of 687 papers associated with the Wellcome Trust published in 2005 found a moderate positive Spearman correlation ( $\rho = 0.445$ ) between the importance of papers, as judged by Wellcome Trust reviewers (each paper was reviewed by two reviewers using a four-point scale), and those identified by the Faculty Opinions experts. The study also found a moderate positive correlation (0.45) between

<sup>34</sup>[https://facultyopinions.com/wp-content/uploads/2020/07/Faculty-Opinions\\_Reference\\_Guide.pdf](https://facultyopinions.com/wp-content/uploads/2020/07/Faculty-Opinions_Reference_Guide.pdf)

Wellcome Trust reviewer ratings and citation counts to articles 3 years after the reviews had been conducted. The Journal Impact Factor had the strongest association (0.625) with Wellcome Trust reviewer ratings, however (Allen et al., 2009).

## Potential for artificial intelligence

Despite the research reviewed above, there do not seem to have been any AI analyses of post-publication peer review. This is presumably because it is scarce and AI needs a large amount of data to work well. If the number of post-publication peer review comments increases sufficiently then it would become possible to develop applications to help estimate the quality of published research through these post-publication comments. In this situation, the above results suggest that there will be disciplinary differences in the number of reviews and therefore probably the accuracy of future programs. Moreover, thought should be given to ethical considerations to avoid penalizing some sectors of society through algorithmic bias due to the gender and geographic imbalances mentioned above.

## CONCLUSIONS

The evidence above conclusively shows that AI is useful for helping to find reviewers and the spread of this technology to other contexts where it is not yet used, such as REF reviewer assignments, is recommended. There is also evidence that AI can sometimes support the initial quality control of submitted manuscripts. Although plagiarism detection is the obvious example, and is presumably widely used by publishers, statistical checking also seems useful and extending the capability of such software would be valuable. In contrast, there is insufficient evidence yet to use AI to support reviewing and it should not be used to replace human reviewers. Further testing of software for reviewing is important, however, perhaps for desk rejection of obviously poor papers. Finally, whilst peer review text and scores are currently too sparsely available to support post-publication research assessment, further research will be helpful if this changes in the future.

## AUTHOR CONTRIBUTIONS

Kayvan Kousha wrote the article. Mike Thelwall revised it.

## ACKNOWLEDGEMENTS

This review was funded by Research England, Scottish Funding Council, Higher Education Funding Council for Wales, and Department for the Economy, Northern Ireland as part of the Future Research Assessment Programme ([www.jisc.ac.uk/future-research-assessment-programme](http://www.jisc.ac.uk/future-research-assessment-programme)). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

No new data was created by the research for this article.

## REFERENCES

- Al Mahmud, T., Hossain, B. M., & Ara, D. (2018). Automatic reviewers assignment to a research paper based on allied references and publications weight. In *2018 4th International Conference on Computing Communication and Automation (ICCCA)* (pp. 1–5). IEEE.
- Allen, L., Jones, C., Dolby, K., Lynn, D., & Walport, M. (2009). Looking for landmarks: The role of expert review and bibliometric analysis in evaluating scientific publication outputs. *PLoS One*, *4*(6), e5910. <https://doi.org/10.1371/journal.pone.0005910>
- American Journal Experts. (2018). Peer review: How we found 15 million hours of lost time. [www.aje.com/en/arc/peer-review-process-15-million-hours-lost-time/](http://www.aje.com/en/arc/peer-review-process-15-million-hours-lost-time/)
- Bornmann, L. (2015). Interrater reliability and convergent validity of F 1000 P rime peer review. *Journal of the Association for Information Science and Technology*, *66*(12), 2415–2426. <https://doi.org/10.1002/asi.23334>
- Bornmann, L., & Leydesdorff, L. (2013). The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from InCites and F1000. *Journal of Informetrics*, *7*(2), 286–291. <https://doi.org/10.1016/j.joi.2012.12.003>
- Bornmann, L., & Leydesdorff, L. (2015). Does quality and content matter for citedness? A comparison with Para-textual factors and over time. *Journal of Informetrics*, *9*(3), 419–429. <https://doi.org/10.1016/j.joi.2015.03.001>
- Burley, R., & Moylan, E. (2017). What might Peer Rev look like in 2030? A report from BioMed Central and Digital Science. [https://figshare.com/articles/journal\\_contribution/What\\_might\\_peer\\_review\\_look\\_like\\_in\\_2030\\_/4884878/1](https://figshare.com/articles/journal_contribution/What_might_peer_review_look_like_in_2030_/4884878/1)
- Cabanac, G., Labbé, C., & Magazinov, A. (2021). Tortured phrases: A dubious writing style emerging in science. Evidence of critical issues affecting established journals. *arXiv preprint arXiv: 2107.06751*.
- Chakraborty, S., Goyal, P., & Mukherjee, A. (2020). Aspect-based sentiment analysis of scientific reviews. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020* (pp. 207–216).
- Charlin, L., & Zemel, R. (2013). The Toronto paper matching system: An automated paper-reviewer assignment system. <https://openreview.net/pdf?id=caynafZAnBafx>
- Checco, A., Bracciale, L., Loreti, P., Pinfield, S., & Bianchi, G. (2021). AI-assisted peer review. *Humanities and Social Sciences Communications*, *8*(1), 1–11. <https://doi.org/10.1057/s41599-020-00703-8>
- Cyranoski, D. (2019). Artificial intelligence is selecting grant reviewers in China. *Nature*, *569*(7756), 316–318. <https://doi.org/10.1038/d41586-019-01517-8>
- Du, J., Tang, X., & Wu, Y. (2016). The effects of research level and article type on the differences between citation metrics and F 1000 recommendations. *Journal of the Association for Information Science and Technology*, *67*(12), 3008–3021. <https://doi.org/10.1002/asi.23548>
- Feng, X., Zhang, H., Ren, Y., Shang, P., Zhu, Y., Liang, Y., & Xu, D. (2019). The deep learning-based recommender system “Pubmender” for choosing a biomedical publication venue: Development and validation study. *Journal of Medical Internet Research*, *21*(5), 27–35. <https://doi.org/10.2196/12957>

- Fiez, T., Shah, N., & Ratliff, L. (2020). A SUPER\* algorithm to optimize paper bidding in peer review. In *Conference on Uncertainty in Artificial Intelligence* (pp. 580–589). PMLR. <http://proceedings.mlr.press/v124/fiez20a/fiez20a.pdf>
- Ghosal, T., Raj, A., Ekbal, A., Saha, S., & Bhattacharyya, P. (2019). A deep multimodal investigation to determine the appropriateness of scholarly submissions. In *In 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 227–236). IEEE Press.
- Ghosal, T., Verma, R., Ekbal, A., & Bhattacharyya, P. (2019). DeepSentiPeer: Harnessing sentiment in review texts to recommend peer review decisions. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 1120–1130).
- Heaven, D. (2018). The age of AI peer reviews. *Nature*, 563(29), 609–610. <https://doi.org/10.1038/d41586-018-07245-9>
- Hosseini, M., & Horbach, S. P. (2023). Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Research Integrity and Peer Review*, 8(1), 4. <https://doi.org/10.1186/s41073-023-00133-5>
- Iana, A., & Paulheim, H. (2021). GraphConfRec: A graph neural network-based conference recommender system. In *2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 90–99). IEEE.
- Kalmukov, Y. (2020). An algorithm for automatic assignment of reviewers to papers. *Scientometrics*, 124(3), 1811–1850. <https://doi.org/10.1007/s11192-020-03519-0>
- Kalnins, A. U., Halm, K., & Castillo, M. (2015). Screening for self-plagiarism in a subspecialty-versus-general imaging journal using iThenticate. *American Journal of Neuroradiology*, 36(6), 1034–1038. <https://doi.org/10.3174/ajnr.A4234>
- Khalil, M., & Er, E. (2023). Will ChatGPT get you caught? Rethinking of plagiarism detection. *arXiv preprint arXiv:2302.04335*.
- Kim, S., Choi, H., Kim, N., Chung, E., & Lee, J. Y. (2018). Comparative analysis of manuscript management systems for scholarly publishing. *Science Editing*, 5(2), 124–134. <https://doi.org/10.6087/kcse.137>
- Kovanis, M., Porcher, R., Ravaud, P., & Trinquart, L. (2016). The global burden of journal peer review in the biomedical literature: Strong imbalance in the collective enterprise. *PLoS One*, 11(11), e0166387. <https://doi.org/10.1371/journal.pone.0166387>
- Li, X., & Thelwall, M. (2012). F1000, Mendeley and traditional bibliometric indicators. *Proceedings of the 17th International Conference on Science and Technology Indicators* (Vol. 2, pp. 451–551).
- Li, X., & Watanabe, T. (2013). Automatic paper-to-reviewer assignment, based on the matching degree of the reviewers. *Procedia Computer Science*, 22, 633–642. <https://doi.org/10.1016/j.procs.2013.09.144>
- Lin, J., Song, J., Zhou, Z., & Shi, X. (2021). Automated scholarly paper review: Possibility and challenges. *arXiv preprint arXiv:2111.07533*.
- Liu, O., Wang, J., Ma, J., & Sun, Y. (2016). An intelligent decision support approach for reviewer assignment in R&D project selection. *Computers in Industry*, 76, 1–10. <https://doi.org/10.1016/j.compind.2015.11.001>
- Liu, X., Wang, X., & Zhu, D. (2022). Reviewer recommendation method for scientific research proposals: A case for NSFC. *Scientometrics*, 127, 3343–3366. <https://doi.org/10.1007/s11192-022-04389-4>
- Lund, B. D., Wang, T., Mannuru, N. R., Nie, B., Shimray, S., & Wang, Z. (2023). ChatGPT and a new academic reality: Artificial intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology*, 74(5), 570–581. <https://doi.org/10.1002/asi.24750>
- Luo, J., Feliciani, T., Reinhart, M., Hartstein, J., Das, V., Alabi, O., & Shankar, K. (2021). Analysing sentiments in peer review reports: Evidence from two science funding agencies. *Quantitative Science Studies*, 2(4), 1271–1295.
- Menke, J., Roelandse, M., Ozyurt, B., Martone, M., & Bandrowski, A. (2020). The rigor and transparency index quality metric for assessing biological and medical science methods. *IScience*, 23(11), 101698. <https://doi.org/10.1016/j.isci.2020.101698>
- Mohammadi, E., & Thelwall, M. (2013). Assessing non-standard article impact using F1000 labels. *Scientometrics*, 97(2), 383–395. <https://doi.org/10.1007/s11192-013-0993-9>
- Mrowinski, M. J., Fronczak, P., Fronczak, A., Ausloos, M., & Nedic, O. (2017). Artificial intelligence in peer review: How can evolutionary computation support journal editors? *PLoS One*, 12(9), e0184711. <https://doi.org/10.1371/journal.pone.0184711>
- Nuijten, M. B., & Polanin, J. R. (2020). “Statcheck”: Automatically detect statistical reporting inconsistencies to increase reproducibility of meta-analyses. *Research Synthesis Methods*, 11(5), 574–579. <https://doi.org/10.1002/jrsm.1408>
- Ortega, J. L. (2017). Are peer-review activities related to reviewer bibliometric performance? A scientometric analysis of Publons. *Scientometrics*, 112(2), 947–962. <https://doi.org/10.1007/s11192-017-2399-6>
- Ortega, J. L. (2019). Exploratory analysis of Publons metrics and their relationship with bibliometric and altmetric impact. *Aslib Journal of Information Management*, 71(1), 124–136. <https://doi.org/10.1108/AJIM-06-2018-0153>
- Ortega, J. L. (2022). Classification and analysis of PubPeer comments: How a web journal club is used. *Journal of the Association for Information Science and Technology*, 73(5), 655–670. <https://doi.org/10.1002/asi.24568>
- Pradhan, T., Bhatia, C., Kumar, P., & Pal, S. (2021). A deep neural architecture based meta-review generation and final decision prediction of a scholarly article. *Neurocomputing*, 428, 218–238. <https://doi.org/10.1016/j.neucom.2020.11.004>
- Pradhan, T., & Pal, S. (2020). CNAVER: A content and network-based academic venue recommender system. *Knowledge-Based Systems*, 189, 105092. <https://doi.org/10.1016/j.knosys.2019.105092>
- Price, S., & Flach, P. A. (2017). Computational support for academic peer review: A perspective from artificial intelligence. *Communications of the ACM*, 60(3), 70–79. <https://doi.org/10.1145/2979672>
- REF. (2022). Guidance on the REF2021 results. <https://ref.ac.uk/guidance-on-results/guidance-on-ref-2021-results/>
- Resnik, D. B., & Elmore, S. A. (2016). Ensuring the quality, fairness, and integrity of journal peer review: A possible role of editors. *Science and Engineering Ethics*, 22, 169–188. <https://doi.org/10.1007/s11948-015-9625-5>
- Roberts, J., & Fisher, D. (2020). pReview: The artificially intelligent conference reviewer. In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 665–668). IEEE.
- Severin, A., Strinzel, M., Egger, M., Domingo, M., & Barros, T. (2021). Characteristics of scholars who review for predatory and legitimate journals: Linkage study of Cabells scholarly analytics and Publons data. *BMJ Open*, 11(7), e050270. <https://doi.org/10.1136/bmjopen-2021-050270>

- Skorikov, M., & Momen, S. (2020). Machine learning approach to predicting the acceptance of academic papers. In *In 2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)* (pp. 113–117). IEEE.
- Smith, Z. L., Chiang, A. L., Bowman, D., & Wallace, M. B. (2019). Longitudinal relationship between social media activity and article citations in the journal *gastrointestinal endoscopy*. *Gastrointestinal Endoscopy*, 90(1), 77–83.
- Tedford, A. (2015). *Rolling out our new editorial system: EVISE®: Find out how the new system will help reviewers streamline their workload*. Elsevier [www.elsevier.com/connect/reviewers-update/rolling-out-our-new-editorial-system-evise](http://www.elsevier.com/connect/reviewers-update/rolling-out-our-new-editorial-system-evise)
- Thelwall, M. (2022). Journal and disciplinary variations in academic open peer review anonymity, outcomes, and length. *Journal of Librarianship and Information Science*, 55, 299–312. <https://doi.org/10.1177/09610006221079345>
- Thelwall, M., Kousha, K., Abdoli, M., Stuart, E., Makita, M., Wilson, P., & Cancellieri, M. (2022). Predicting article quality scores with machine learning: The UK Research Excellence Framework. *arXiv preprint arXiv:2212.05415*.
- Thelwall, M., Papas, E. R., Nyakoojo, Z., Allen, L., & Weigert, V. (2020). Automatically detecting open academic review praise and criticism. *Online Information Review*, 44(5), 1057–1076. <https://doi.org/10.1108/OIR-11-2019-0347>
- Waltman, L., & Costas, R. (2014). F 1000 recommendations as a potential new data source for research evaluation: A comparison with citations. *Journal of the Association for Information Science and Technology*, 65(3), 433–445. <https://doi.org/10.1002/asi.23040>
- Wang, D., Liang, Y., Xu, D., Feng, X., & Guan, R. (2018). A content-based recommender system for computer science publications. *Knowledge-Based Systems*, 157, 1–9. <https://doi.org/10.1016/j.knsys.2018.05.001>
- Wang, K., & Wan, X. (2018). Sentiment analysis of peer review texts for scholarly papers. In the 41st international ACM SIGIR Conference on Research & Development in Information Retrieval (pp. 175–184).
- Wang, P., & Su, J. (2021). Post-publication expert recommendations in faculty opinions (F1000Prime): Recommended articles and citations. *Journal of Informetrics*, 15(3), 101174. <https://doi.org/10.1016/j.joi.2021.101174>
- Wardle, D. A. (2010). Do Faculty of 1000' (F1000) ratings of ecological publications serve as reasonable predictors of their future impact? *Ideas in Ecology and Evolution*, 3, 11–15. <https://doi.org/10.4033/iee.2010.3.3.c>
- Wilkinson, J., & Down, P. (2018). Publons: Releasing the untapped power of peer review for universities. *Insight*, 31, 20. <https://doi.org/10.1629/uksg.407/>
- Wolfram, D., Wang, P., Hembree, A., & Park, H. (2020). Open peer review: Promoting transparency in open science. *Scientometrics*, 125(2), 1033–1051. <https://doi.org/10.1007/s11192-020-03488-4>
- Yuan, W., Liu, P., & Neubig, G. (2021). Can we automate scientific reviewing? *arXiv preprint arXiv:2102.00176*.
- Zaharie, M. A., & Osoian, C. L. (2016). Peer review motivation frames: A qualitative approach. *European Management Journal*, 34(1), 69–79.
- ZhengWei, H., JinTao, M., YanNi, Y., Jin, H., & Ye, T. (2022). Recommendation method for academic journal submission based on doc2vec and XGBoost. *Scientometrics*, 127, 2381–2394. <https://doi.org/10.1007/s11192-022-04354-1>
- Zong, Q., Fan, L., Xie, Y., & Huang, J. (2020). The relationship of polarity of post-publication peer review to citation count: Evidence from Publons. *Online Information Review*, 44(3), 583–602. <https://doi.org/10.1108/OIR-01-2019-0027>