AI REVIEWER WITH AN OPEN PROMPT: A UNIVERSAL PROMPT FOR HYBRID EXPERTISE AND SELF-ASSESSMENT OF SCIENTIFIC TEXTS

Kravtsov Gennadiy Grigorievich - Director of the Research Center "Applied Statistics". ORCID: https://orcid.org/0009-0000-3405-1461

Abstract

The article explores overcoming the crisis of scientific peer review through a hybrid model that combines algorithmic analysis with expert evaluation. The key contribution is an open universal prompt for AI-based review, tested on a corpus of scientific texts. The prompt is intended for widespread use by authors for preliminary self-assessment and is published in full with the expectation of its broad application and collective improvement by the community. An innovative publication approach is demonstrated, including an AI review and raw data. It is argued that hybrid peer review and AI self-analysis are not merely a technological trend but a necessary condition for the advancement of science in the digital age.

Keywords: artificial intelligence (AI), scientific peer review, hybrid peer-review, open science, prompt engineering, AI assistant, manuscript self-assessment, objective evaluation, algorithmic assessment, preprint culture, collective improvement, AMKO, free AI tool, thesis checking

Introduction

Modern scientific communication faces a paradox: while the number of publications has surged dramatically, their quality has been steadily declining [Bauchner & Rivara, 2024]. Of particular concern is the flood of formalistic papers created solely to meet quantitative metrics, as well as the stream of AI-generated content that often merely reiterates known concepts without scientific novelty.

In this context, the integration of artificial intelligence into the peer review process is transitioning from an experimental practice to an urgent necessity. Furthermore, the very feasibility of using AI in writing scientific articles is no longer in doubt [Bauchner & Rivara, 2024].

However, existing solutions are often fragmented or commercial. In response to this challenge, this article not only proposes a theoretical model of hybrid expertise but

also presents a practical tool—an open, universal prompt for AI-assisted reviewing. This prompt, tested on a significant dataset, allows every author to conduct a preliminary analysis of their work against standardized criteria. We are publishing it in its entirety, anticipating its widespread adoption and further evolution by the scientific community.

Modern artificial intelligence is fundamentally incapable of generating ideas in the human sense of the process. Its core function is the processing and recombination of existing data to generate content based on it. The key role of the human is to formulate the initial creative impulse—the prompt—which the AI then develops, structures, and articulates into a concrete form. Thus, AI acts not as a source but as an executor, transforming the author's intent into a detailed and formatted result, yet always within the conceptual boundaries set by the human.

The key to effective application of this technology lies in the optimal division of functions between the researcher and the algorithm. The human retains creative and managerial tasks: conceptualization and overall control [Rennie, 2003]. Meanwhile, artificial intelligence assumes an instrumental role: formulating clear and literate text [Cherkasova et al., 2024], consulting on emerging questions, and searching for and formatting supporting material [Budaeva et al., 2024].

This article proposes to expand the existing model of AI application by considering AI not only as an assistant tool for writing scientific texts but also as a powerful means for conducting preliminary expert assessment. This approach aims to identify and correct a paper's weaknesses in advance, before its formal submission to a scientific journal, thereby enhancing publication quality and the efficiency of the review process.

As Yuri Medvedev (2023) notes, AI has already demonstrated significant potential for the structured and impartial analysis of scientific texts. Notably, according to a survey, 82% of experts acknowledged the comparability of algorithmic assessment to traditional expertise in terms of objectivity and depth of analysis. However, despite these advantages, the use of AI in reviewing remains unsystematic.

The scientific novelty of this work is as follows:

1. A model of hybrid reviewing has been developed, classifying manuscripts by their degree of novelty to optimize the allocation of resources between AI and human experts.

2. An open, universal prompt for the initial review stage has been created and tested. It standardizes article evaluation against key criteria and is suitable for author self-assessment (see the file Universal_AI_Review_Prompt_v_31_EN.txt in the supplementary materials) .

3. A "open expertise" publication model has been proposed, which includes, alongside the text and data, an AI-generated review (see the file AI_Review_Open_Source_Article_EN.pdf in the supplementary materials), thereby enhancing transparency and verifiability.

The proposed approach optimizes the workload of experts, reduces reviewer time, significantly decreases the subjectivity of evaluations, and helps authors identify weaknesses in their work.

Of course. Here is the English translation of the provided text.

The Hybrid System of Scientific Peer Review

The modern institution of scientific peer review contains a fundamental contradiction: intended to be an arbiter of novelty [Menke et al., 2022], it often itself becomes a factor hindering the development of science. Reviewers formed within the framework of traditional views often exhibit conservatism when evaluating breakthrough approaches [Haffar et al., 2019]. This initially protective mechanism of the scientific community, in new realities, risks becoming a certain barrier to scientific progress.

The technological revolution of recent decades could not but affect the sphere of scientific expertise. Artificial intelligence already demonstrates impressive capabilities in several key areas: processing large datasets, identifying cases of plagiarism, and formally checking articles for compliance with requirements. However, in analyzing complex graphical dependencies, interpreting non-standard results, or assessing methodological novelty, algorithms still fall short of qualified human experts [Freyer & Ryan, 2024].

A promising solution to this dilemma could be the practice of publishing raw data, allowing for independent verification of results [Bauchner & Rivara, 2024]. This approach was implemented in the monograph on the AMKO method [Kravtsov, 2025b], where the author provided not only the research text but also a full set of accompanying materials [Kravtsov, 2025c].

Currently, a hybrid model of scientific peer review seems optimal, implementing a multi-level approach where artificial intelligence and human expertise do not compete but effectively complement each other (Fig. 1).

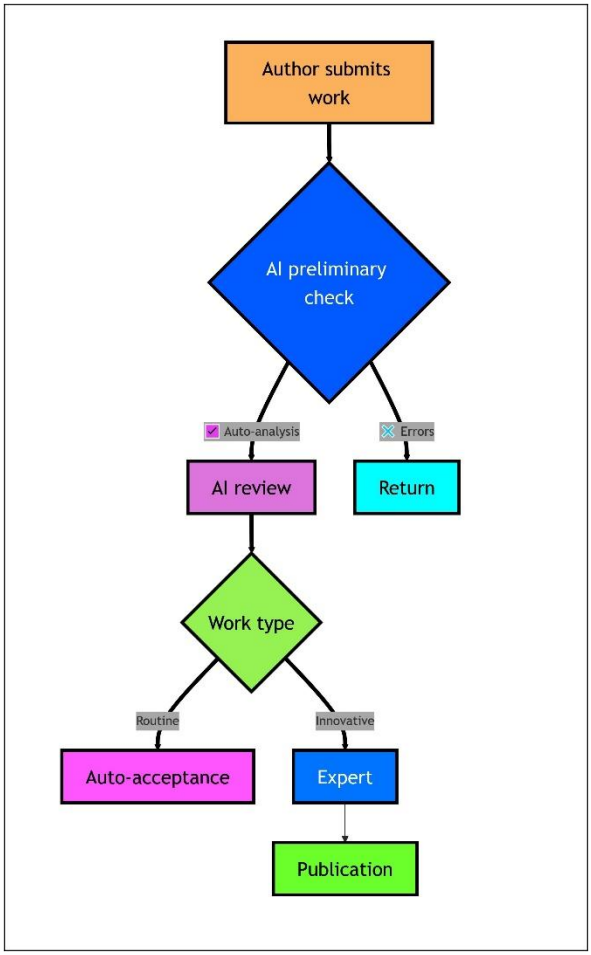

Figure 1. Flowchart of a hybrid review system integrating AI and expert assessment

At the first stage, AI conducts a preliminary analysis, checking compliance with formal requirements and the technical correctness of the presented data. After successful initial verification, the system classifies the work by its degree of novelty: standard studies with established methodology may receive an automated review, while truly innovative works are sent to experts for in-depth analysis.

The main advantage of this approach lies in the rational distribution of resources: routine checks are automated, allowing experts to focus on analyzing the most complex and innovative works. As algorithms develop, the model can adapt, maintaining an optimal balance between automation and expert assessment.

Advantages of AI Reviewing

Modern reviewing algorithms provide a qualitatively new level of scientific expertise due to their undeniable advantages.

1. Independence of Assessment

Algorithmic systems eliminate the influence of subjective factors that traditionally distort scientific evaluation. Unlike human reviewers, AI is not susceptible to the authority of researchers, their institutional affiliation, or geographical origin. This

creates fundamentally equal conditions for all scientific works, regardless of their origin [Bauchner & Rivara, 2024; Haffar et al., 2019].

2. High Analysis Speed

AI systems can process materials in seconds, whereas traditional reviewing takes weeks or months. Such efficiency is especially important in rapidly developing scientific fields where the timeliness of publication is critical.

3. Identification of Hidden Patterns

Thanks to their ability to work with big data, algorithms detect interconnections and trends that often escape human attention. This allows for identifying promising research directions and potential methodological problems.

4. Universal Evaluation Standards

The AI expertise system uses a unified 100-point scale to assess the quality of the provided material, considering five key parameters: scientific novelty (30%), methodological rigor (25%), practical value (20%), quality of data visualization (15%), and ethical aspects (10%). This approach ensures comparability of assessments [Bauchner et al., 2024, Menke et al., 2022].

5. Integrated Expertise

The platform combines automated processing with expert involvement. AI performs initial analysis and checks formal criteria, leaving substantive evaluation and interpretation of results to specialists. Simultaneously, the system tracks citations and the actual impact of publications, adjusting their rating.

6. Economic Component

The implementation of AI reviewing ensures significant economic efficiency of the scientific publishing process by reducing operational costs for expert labor and administrative overhead, as algorithms can handle up to 90% of routine checks. Consequently, it becomes possible to reduce publication costs.

Such an architecture of scientific expertise creates an optimal balance between technological efficiency, academic rigor, and economic feasibility. Its implementation will accelerate scientific progress while ensuring high standards of quality and objectivity.

Modern AI Systems for Scientific Reviewing: A Comparative Analysis

1. **ScholarOne Platform (Clarivate)** — An online platform for managing the process of submission, review, and publication of scientific articles. It automates interaction between authors, reviewers, and editors, supports various review models, and integrates with other Silverchair services [Silverchair, 2025].

2. **OpenReview Platform** — An online service for open and transparent peer review of scientific publications. It allows conference organizers, journals, and other scientific communities to flexibly configure expert evaluation processes, supporting various levels of openness and discussion. OpenReview maintains a history of all review stages, including feedback, comments, and dialogue between authors and reviewers, promoting open access to materials and scientific discourse [OpenReview, 2025].

3. **SciScore System** — An automated tool for checking the methodology and reporting of scientific research. It analyzes the "Materials and Methods" sections in articles to assess compliance with principles of rigor and transparency, helps improve research reproducibility by generating a report with scores for data quality and identification of key resources (antibodies, cell lines, plasmids, etc.). According to Menke et al. (2022), its algorithms match expert assessment in 89% of cases. A critical limitation is its narrow specialization: the system is not adapted for social sciences [SciScore, 2025].

4. **IBM Watson's Peer Review Assistant Tool** — An artificial intelligence created to assist in reviewing scientific articles. It analyzes scientific texts, helps identify key points, possible errors, and inconsistencies, and speeds up and improves the process of preparing reviews for research journals. The assistant uses natural language processing and machine learning technologies to enhance the quality and efficiency of expert evaluation [IBM, 2025].

Practical Application of AI Reviewing

The implementation of artificial intelligence in the scientific evaluation process is naturally transitioning from the theoretical to the practical plane. This research aimed to create not just a concept, but a working tool—a universal prompt for AI reviewing, suitable for broad application (see the file Universal_AI_Review_Prompt_v_31_EN.txt in the supplementary materials).

The prompt was developed on the DeepSeek platform and represented an iterative process of collaboration between the researcher and the algorithm. The source material was a basic set of criteria for evaluating a scientific article, formulated by the author. Then, through dialogue, step-by-step optimization was carried out: clarifying wording, making them more specific, structuring, and adapting them to the research goals. A simple request to write a review does not yield a satisfactory result; the key to success lies in a detailed and refined instruction.

During development, the following problems of the initial prompt versions were identified and solved:

1. Result Instability. When re-analyzing the same article, the AI could produce different assessments and formulations due to its probabilistic nature. This problem was mitigated by tightening the query structure, minimizing the arbitrariness of

interpretations. Variability within acceptable limits can now be considered analogous to differences between the opinions of several human reviewers.

2. Focus on Formal Templates. The algorithm often criticized the absence of elements it expected (e.g., specific statistical tests) without evaluating the validity of the author's methodological choice. The solution was to include in the prompt requirements to analyze not only the presence but also the adequacy of the methods used.

3. Lack of an Absolute Quality Benchmark. The AI's assessment of a work is contextual. To calibrate the system, it proved useful to simultaneously provide several articles on a similar topic, allowing the algorithm to form a more objective comparative understanding.

4. Skipping Mandatory Criteria. Some prompt points were occasionally ignored. This shortcoming has been largely eliminated through rigid structuring of the query with a clear sequence for performing each analysis step.

The resulting universal prompt was tested on a representative sample comprising over 50 scientific articles and reports from various disciplinary fields. Ethically, we do not have the right to publish reviews of others' materials without the authors' permission. Therefore, to demonstrate the method's operation, a self-check approach was applied: the prompt was used to analyze this and other articles of varying levels by the author (see the files AI_Review_Open_Source_Article_EN.pdf, Comparative_Table_4_Articles_Quality.pdf in the supplementary materials). The AI-generated review, raw data, and accompanying materials are published in open access in the Dataverse repository, forming a single verifiable research complex.

The most indicative example of the practical application of the method was the AI review of the monograph "Introduction to the Analytical Method for Monitoring Education (AMKO)" [Kravtsov, 2025b]. This process proved extremely valuable for authorial self-analysis. Despite a high overall rating, the system identified substantive weaknesses previously overlooked.

For instance, the algorithm pointed to insufficient clarity in describing the basic principle of the method—using data from the period 1995–2005 as a fixed reference point. Interpreting this data as "outdated," the AI reviewer recommended including more recent results, thereby demonstrating a classic misunderstanding of the axiomatic nature of a reference point, the immutability of which is a condition for maintaining data comparability in the long term. A similar question may arise from human reviewers.

Furthermore, questions were raised about the reasons for choosing the subject of history to illustrate the method, as well as about the insufficient, in the algorithm's opinion, completeness of the description of computational procedures. A key point was that the AI did not recognize that the provided examples were a demonstration of a

large-scale algorithm, the full description of which requires a separate publication cycle. Criticism related to suggestions for a large-scale comparative analysis with international studies (PISA), although useful as a perspective, often exceeds the resource capabilities of an individual study.

This dialogue with the AI reviewer allowed the author to more precisely define the methodological boundaries of their research, further justify key assumptions, and ultimately improve the clarity and validity of the text.

To work with the prompt, the following procedure is recommended: open a new chat in DeepSeek, upload the prompt text, then provide the article for analysis and a detailed description of additional materials. To reduce random fluctuations in results, it is useful to run the same article several times and consider the averaged outcome. If any prompt points are skipped, point this out and repeat the request. For calibration, it is recommended to use a similar request for several articles, preferably of different levels— at least three. Usually, articles from the reference list are suitable for this. After which it is advisable to ask the system: "output a comparative analysis of all articles according to the specified criteria in the form of a summary table." The prompt is developed in Russian and English, and the system processes articles in both languages equally correctly.

One of the most valuable aspects of the tool is its ability to conduct a comparative analysis of several works. When loaded simultaneously, the algorithm systematizes the data and provides structured comparisons, including tables with key parameters of the studies (Fig. 2). This allows the author to quickly identify gaps in their own work and address them before submitting the article to a journal.

Criterion	AMKO	Teaching Quality (...)	Point-Rating System
Methodology	Mathematical modeling (20-point scale, polynomial regression)	Psychometric analysis (CFA), questionnaire adaptation	Theoretical analysis, thought experiments
Sample	1,651 records (history, 10th grade) + 2.5 million records in educational database	N students	No empirical data
Data Type	Objective educational metrics + Excel tables	Subjective student evaluations (questionnaires)	Theoretical reasoning
Verifiability	Full (open source data)	Partial (methods described, but no raw data)	Impossible (no data)
Novelty	Fundamentally new assessment system	Adaptation of foreign methodology for Russian context	Critical analysis of existing system
Practical Value	Ready-to-use tool for schools with AI integration	Recommendations for teacher professional development	Proposals for modifying point-rating system
Strengths	• Objective metrics • Open data • Machine processing	• Large sample • Validated tool • Multidimensional analysis	• Systemic problem analysis • Practical recommendations
Weaknesses	Limited to one subject (history)	• Subjectivity of student evaluations • No raw data provided	Lack of empirical testing
Data Visualization	Dependency graphs, transformation tables	Correlation and regression tables	None
Recommendations	Comparison with international analogues	Publication of anonymized data	Pilot testing of proposals

Figure 2. Comparative table of methodologies (raw data of the two compared methodologies altered for ethical reasons)

Usage Recommendations:

1. For Authors:

* Use for preliminary self-check.

* It is recommended to analyze several articles on the same topic simultaneously for comparative calibration.

* Critically evaluate AI comments regarding methodological specifics.

2. For Editorial Boards:

* Can serve as a primary filter for reviewers.

* Allows standardization of formal evaluation criteria.

* Reduces the workload on human resources by 40-60%, and for some review-analytical works by up to 90%.

3. For Research Supervisors:

* A tool for teaching graduate students the principles of scientific criticism.

* Possibility for objective assessment of intermediate results.

The proposed open prompt serves as an assistant tool for authors, helping to improve publication quality, ensure compliance with academic standards, and conduct preliminary self-analysis, thereby reducing the load on traditional reviewers. However, it should be considered as a basis for further development and improvement, which is what is proposed to users.

Conclusion

The integration of artificial intelligence into the process of authorial control and reviewing represents a natural stage in the development of academic culture [Budaeva, Zyryanova, 2024]. The example of including an AI review and accompanying raw materials as part of a publication creates a fundamentally new model of "open expertise," which paves the way for a fairer and more efficient system of scientific communication.

The key practical result of this work is an open universal prompt that can provide real help to numerous authors today. We view it not as a frozen dogma, but as a starting point for collective creativity. Its publication invites all interested researchers to test, critique, and jointly improve this tool.

For successful AI integration, a number of tasks need to be solved: developing clear ethical standards [Ivanova, 2024], creating mechanisms for verifying algorithmic decisions, and preserving the key role of the expert in evaluating fundamentally new approaches. As practice shows, scientists who effectively use AI gain a serious advantage, but technology should not replace, but rather enhance, traditional expertise.

We believe that in the coming years, a stable hybrid review system will form, where routine operations will be delegated to algorithms, and strategic issues will remain with the scientific community. This symbiosis opens the way to creating a fairer, more efficient, and transparent system of scientific communication that meets the demands of the digital age.

Reference List

1. Bauchner, H., & Rivara, F. P. (2024). Use of artificial intelligence and the future of peer review. *Health Affairs Scholar, 2*(5), qxae058. https://doi.org/10.1093/haschl/qxae058

2. Budaeva, D. A., & Zyryanova, I. N. (2024). [Pragmalinguistic aspects of formulating prompts for ChatGPT in scientific communication]. *Kazanskaya Nauka [Kazan Science]*, 7, 220-224. EDN: NAUXGT. (in Russian)

3. Freyer, F. J., & Ryan, A. (2024, March 20). How does bad data slip through? Allegations of research fraud raise questions about "peer review". *The Boston Globe*.

4. Haffar, S., Bazerbachi, F., & Murad, M. H. (2019). Peer review bias: A critical review. *Mayo Clinic Proceedings, 94*(4), 670–676. https://doi.org/10.1016/j.mayocp.2018.09.004

5. IBM. (n.d.). *Watsonx Assistant*. Retrieved August 21, 2024, from https://www.ibm.com/products/watsonx-assistant

6. Ivanova, L. A. (2024). [Artificial intelligence in writing scientific articles - a positive or harmful factor?]. *Crede Experto: Transport, Obshchestvo, Obrazovanie, Yazyk [Crede Experto: Transport, Society, Education, Language]*, 4(43). https://doi.org/10.51955/2312-1327_2024_4_6 (in Russian)

7. Kravtsov, G. G. (2025a). AI-reviewer with open code: A universal prompt for hybrid expertise and self-analysis of scientific texts [Software; Dataset]. Zenodo. https://doi.org/10.5281/zenodo.17020101 (in Russian)

8. Kravtsov, G. G. (2025b). [Introduction to the Analytical Method of Education Control (AMEC): From mathematical analytics of the educational process to a standardized educational database]. SRC "Applied Statistics". https://doi.org/10.5281/zenodo.16791743 (in Russian)

9. Kravtsov, G. (2025c). *Supplementary Materials for: "Introduction to the Analytical Method of Education Control (AMEC): From Mathematical Analysis of Learning Processes to a Standardized Educational Database"* [Data set]. Harvard Dataverse. https://doi.org/10.7910/DVN/DEHUMS

10. Medvedev, Yu. (2023, October 23). [Artificial intelligence successfully reviews scientific articles]. *Rossiyskaya Gazeta [Russian Gazette]*. https://rg.ru/2023/10/23/iskusstvennyj-intellekt-uspeshno-recenziruet-nauchnye-stati.html (in Russian)

11. Menke, J., Eckmann, P., Ozyurt, I. B., Roelandse, M., Anderson, N., Grethe, J., Gamst, A., & Bandrowski, A. (2022). Establishing institutional scores with the Rigor and Transparency Index: Large-scale analysis of scientific reporting quality. *Journal of Medical Internet Research, 24*(6), e37324. https://www.jmir.org/2022/6/e37324/

12. OpenReview. (n.d.). *Home*. Retrieved August 21, 2024, from https://openreview.net/

13. Rennie, D. (2003). Editorial peer review: Its development and rationale. In F. Godlee & T. Jefferson (Eds.), *Peer review in health sciences* (2nd ed., pp. 1-13). BMJ Books.

14. SciScore. (n.d.). *Home*. Retrieved August 21, 2024, from https://sciscore.com/

15. Silverchair. (n.d.). *Products*. Retrieved August 20, 2024, from https://www.silverchair.com/products/

16. Cherkasova, M. N., & Taktarova, A. V. (2024). [Artificially generated academic text (linguopragmatic aspect)]. *Filologicheskie Nauki. Voprosy Teorii i Praktiki [Philological Sciences. Issues of Theory and Practice]*, 17(7), 2551–2557. https://doi.org/10.30853/phil20240363 (in Russian)