Overview of the AI Review System

The AAAI 2026 AI review system uses a large frontier reasoning model made available through an in-kind sponsorship from OpenAI, along with search and information retrieval, in a multi-step workflow with custom tools to carefully analyse the paper in accordance with the AAAI review criteria before generating the review. Details on the interactions between the AI review and the human reviewers at every stage is detailed in the AAAI 2026 Reviewer Instructions page.

Creation of reviews is only part of the process that carefully examines each submission. There are additional checks, both automated and semiautomated, that check for a variety of potential technical and ethical violations.

The workflow consists of several stages, with each stage designed to focus on different aspects of the paper. The workflow stages include:

- Paper pre-processing: During this step, all paper PDFs are resampled to convert images (if present) to a fixed and consistent resolution, and also converted into a structured markdown using a machine learning based OCR specially trained on scientific PDFs. Both forms of the paper PDF --- the tokenized raw file, and the markdown, are used in subsequent stages. The markdown version in particular extracts math notation and equations into latex format, and extracts table layouts.
- A preliminary analysis of the story of the paper: the problem being addressed, the stated limitations of the state of the art, the key innovations in the paper, and the proofs or empirical evaluations that support the conclusions of the paper.
- Technical accuracy checking: This stage checks for factual, mathematical, and algorithmic correctness, including the equations, pseudocode, and figures in the technical presentation. A code interpreter is available for use by the LLM to assist in this stage.
- Literature search: This stage performs several rounds of literature search of published papers using keyword search based on the above and results from previous rounds of literature search. It uses the retrieved results to examine the contributions of the paper.
- Results checking: This stage checks the results section, including all figures, tables, and text descriptions
- Self-critique: This stage reviews the review itself to check for factual accuracy, consistency with the paper, and any unsubstantiated claims. Revisions are made to address issues found.

The workflow has been developed and refined through extensive testing in several ways, including:

- Testing different workflows on publicly accessible papers and reviews from recent AI conferences
- Testing against specially crafted "test case" papers to test for specific steps of the workflow, including the ability to find mathematical or logical errors
- Multiple sampling with the same inputs to check for consistency across generations
- Testing for robustness to adversarial inputs that may be present in the papers

We will solicit feedback from authors, reviewers, senior program committee, and area chairs to evaluate the quality and helpfulness of the reviews, and to evaluate the pilot AI review program.

The workflows and tests have been informed by findings from previous experiments, tests, and reports on AI systems for scientific review. The list below includes a selection of the articles that have informed the development of this process:

- Can large language models provide useful feedback on research papers? A large-scale empirical analysis.
  https://arxiv.org/pdf/2310.01783
- Is LLM a Reliable Reviewer? A Comprehensive Evaluation of LLM on Automatic Paper Reviewing Tasks
  https://aclanthology.org/2024.lrec-main.816/
- Peer Reviews of Peer Reviews: A Randomized Controlled Trial and Other Experiments
  https://arxiv.org/pdf/2311.09497
- ReviewerGPT? An Exploratory Study on Using Large Language Models for Paper Reviewing
  https://arxiv.org/pdf/2306.00622
- Can LLM feedback enhance review quality? A randomized study of 20K reviews at ICLR 2025
  https://arxiv.org/pdf/2504.09737
- The AI Imperative: Scaling High-Quality Peer Review in Machine Learning
  https://arxiv.org/pdf/2506.08134
- Reviewing Scientific Papers for Critical Problems With Reasoning LLMs: Baseline Approaches and Automatic Evaluation
  https://arxiv.org/pdf/2505.23824
- The Black Spatula Project
  https://the-black-spatula-project.github.io/
- olmOCR: Unlocking Trillions of Tokens in PDFs with Vision Language Models
  https://olmocr.allenai.org/papers/olmocr.pdf

Frequently Asked Questions

1. Q: Which AI models are being used in the AAA-26 AI-Powered Peer Review Assessment System?
   A: The AI review generation uses a large frontier AI model from OpenAI with reasoning capabilities, coupled with several custom tools that the model can decide which, when, and how to use, in a multi-turn workflow.

2. Q: Can authors submit their papers to the AI review system before submission to estimate their chances of acceptance?
   A: No, the authors cannot submit their papers to the AI system before submission, and the review from the AI system will neither estimate chances of acceptance, nor provide any recommendations or scores. Decisions rest solely on humans - the PC members, SPCs, and ACs.

3. Q: If AAAI-26 submits my paper to AI system(s), won't that violate the "unjustified disclosure of information" prescribed by the AAAI Code of Professional Ethics and Conduct?

A: No. AAAI-26 will only utilize AI systems via APIs whose provider guarantees that queries to it , including the submission and our prompts, remain private and will neither be used to train future models, nor used for any other purpose except for explicitly providing outputs for the AI review

4. Q: Who is responsible when an AI-assisted review is wrong or harmful?
   A: In the AAAI-26 pilot program of an AI-assisted peer-review, all decisions are left to humans. Furthermore, the SPC members are responsible for excluding any content that can be considered harmful.

5. Q: How is the use of AI tools in the AAAI-26 reviewing process consistent with privacy protections for authors?
   A: The only information available to the AI review workflow is the submitted anonymous paper PDF. As stated in the author instructions, paper submissions must be anonymous, and authors must take appropriate steps to ensure that their identity is not revealed.

6. Q: How will the committee address the risk of amplifying linguistic, geographic or topical biases baked into the LLMs training data?
   A: The AI-generated review, just like the human reviews, will be inspected by the SPCs to flag ethical violations or inappropriate content, including statements that indicate bias.

7. To what extent will LLMs be used throughout the review workflow, and how will their roles intersect with human reviewers at each decision stage (desk-reject, full review, meta-review, final decision)?
   A: A single AI review will be included in Phase 1, but will not include any scores or recommendations. No human reviewers are replaced, and decisions are made entirely by humans. In phase 2, after the discussion phase, an AI summary of the reviews and points of discussion will be made available to the SPC. Details on the interactions between the AI review and the human reviewers at every stage is detailed in the AAAI 2026 Reviewer Instructions page.

8. Will each submission receive input from multiple AI models, and how will conflicts between outputs—or between AI and human reviewers—be identified and resolved?
   A: The AI review is generated over multiple steps, which include different models, tools, and prompts at each stage, but is synthesized into a single final review. As in the normal review process, there will likely be differences between reviews, whether between different human reviews, or between the AI review and a human review. The rebuttal and discussion phase is intended to address these differences, and allow the human reviewers to arrive at a consensus.

9. What explicit criteria, scoring rubrics, and metrics will the AI systems follow, and how do these align with traditional human review standards?
   A: The AI review will NOT provide any scores, ratings, or recommendations. It is intended to be a factual review, following the same guidelines given to human reviewers.

10. How will AI systems handle submissions that include cutting-edge concepts, controversial findings, niche subfields, complex proofs, specialized notation, or extensive domain-specific terminology?

A: The AI review system parses both the raw submission and a structured markdown version to verify complex proofs, specialized notation, and familiar domain terminology, but they do not make subjective judgments — those remain with human reviewers. The human reviewers are responsible for making subjective judgments based on their understanding, experience, and all the information available to them. As with all AI systems, they have inherent limitations in context retention and domain coverage and are meant solely as an aid to reviewers, SPCs, and ACs, who retain final decision-making authority in their recommendations.

11. How will figures, tables, code, multimedia, and other supplementary materials be processed and evaluated by AI systems?
    A: The AI system uses multimodal LLMs that are capable of interpreting both image, as well as graphical content. To promote high accuracy, in addition to tokenizing the raw paper PDFs, the paper PDFs are also converted using a machine-learning based OCR tool into a structured markdown, preserving table layouts, math notation, and equations. The review generation system uses these multiple formats to interpret figures, tables, code, and other material included in the paper submission.

12. Will the AI system verify references, detect plagiarism, and flag ethical issues in submissions?

    A: The AI review system is not intended to check for plagiarism, and though it utilizes a literature search tool, the use of the tool is primarily to retrieve details on existing related work, not to explicitly check references. Ethical issues may be flagged by the review, but again, are not explicitly intended to be checked by the AI review system. However, there are separate additional checks beyond the AI review system, both automated and semi-automated, in the AAAI review process, which are explicitly designed to check for such violations.

13. Are there length, formatting, keyword, or writing-style guidelines authors should follow to ensure compatibility with AI processing?
    A: No special steps need to be taken by authors, beyond ensuring that the papers follow the AAAI-26 paper formatting guidelines.

14. How are manuscripts stored and processed (on-premises vs. cloud), what contractual safeguards (NDA, GDPR, institutional IP clauses) apply, and can the model or vendor retain or learn from submitted content?
    A: Manuscripts are stored only on OpenReview, and an ephemeral copy is retrieved for the purpose of AI review. Vendor contracts guarantee that data will not be stored or logged, and will not be used for any other purpose beyond generation of the AI review.

15. How will hidden prompts, data poisoning, or other adversarial attacks within manuscripts be detected and mitigated?
    A: The AI system has explicit safeguards (both AI- and classical approach based) in place to detect and flag adversarial attacks, and the system has been tested against known adversarial attacks.

16. What form of feedback will authors receive (raw AI system output, edited summaries, reviewer-approved text), and will it include transparent explanations with highlighted evidence?
    A: Authors will receive the review generated by the AI review system. The review will include

specifics to justify or substantiate any points raised.

17. Will reviews clearly indicate which portions were generated by AI systems versus humans, and who is ultimately accountable for the content?
    A: Yes, the AI review will be clearly marked. Humans are solely responsible for all decision-making in the review process, using all information at their disposal, including the human and AI reviews, the author rebuttals, and the reviewer discussions.

18. What formal mechanisms allow authors to appeal or rebut AI-generated evaluations they believe are incorrect, biased, or incomplete?
    A: Authors may submit a rebuttal to the AI review just like they do for a human review. While the AI review will not be updated based on the authors' response, the human reviewers will be able to read the response, and take it into consideration appropriately during the discussion phase. Major issues that warrant oversight from the conference staff can be flagged using the "comment" field, which is visible only to SPCs and above.

19. How will AI integration affect reviewer workload, diversity, and inclusivity, and may reviewers opt out of using AI assistance?
    A: The AI review system does not replace any human in the process. Authors, reviewers, SPCs, and ACs will have the opportunity to rate and comment on the AI reviews if they wish, to provide feedback on the process.

20. Will AI processing alter review timelines, decision deadlines, or impose additional costs on authors or organizers?
    A: AI processing will not affect the review timelines or decision timelines already established for AAAI 2026.

21. If the underlying model requires updates or patches mid-cycle, how will consistency across previously reviewed papers be ensured?
    A: The AI review generation algorithm will not be updated mid-cycle. The deployment system is designed to be interruptible and resumable to overcome system issues such as network downtime, or server load management.

22. How are the environmental and financial costs of large-scale AI inference being managed?
    A: The financial costs of the AI systems will be cumulatively logged, and reported in a technical report. The environmental impact will be estimated using the latest accepted and published methods based on model usage and compute used.

23. How will the conference measure, report, and iterate on the effectiveness and community impact of AI-assisted reviewing (e.g., experimental design, post-mortem reports, future plans)?
    A: A technical report, summarizing the design, deployment, feedback, impact, and qualitative findings will be written after the event and made available to the community. The report will include findings based on review questionnaires soliciting feedback from paper authors and human reviewers.