

The AAAI-2026 AI-Assisted Peer Review Pilot Program

Joydeep Biswas, Sheila Schoepp, Gautham Vasan, Anthony Opipari, Arthur Zhang, Zichao Hu
Chad Jenkins, Matt Taylor, Peter Stone, Kiri Wagstaff, Matt Lease

Panel and Q&A

Joydeep Biswas
The University of Texas at Austin
Associate Program Chair

Odest Chadwicke Jenkins
University of Michigan
Program Co-Chair

Matthew E. Taylor
University of Alberta
Program Co-Chair

Kevin Leyton-Brown
University of British Columbia
Conference Committee Chair

AI and Program Reviewing



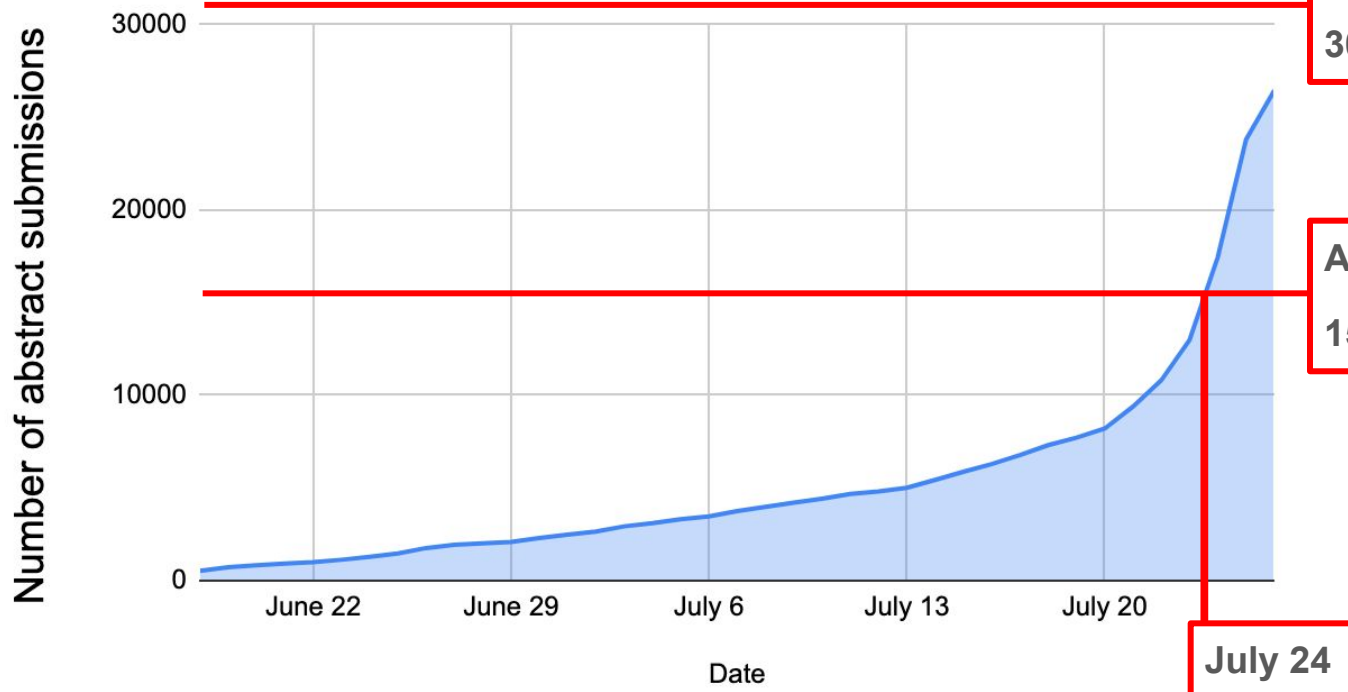
**When you appoint a group of AI experts to run a conference,
and then that conference experiences unprecedented growth
that makes it impossible to continue with business as usual,
what do they do?**

Build a bunch of custom AI tools, of course.

AAAI-26 is the largest AI conference (by number of submissions)



AAAI-26 Main Track Submissions



AAAI-26

30,948 total submissions

AAAI-25

15,532 total submissions

July 24

Core Questions: Review and AI



- How can we do a good job for reviewing with so many papers?

Now and in the future?

- How can we build processes that will continue to do a good job as growth continues?
- What can and should be the role of AI tools?

Review feedback, Scholarly integrity, Paper-reviewer matching ...

Q&A – Post your questions to Slido



AAAI-26

Main Track

Review Process



AAAI-26 Program Team



Program Chairs



Chad
Jenkins



Matthew
Taylor



Bo An



Joydeep
Biswas



David
Crandall



Matthew
Lease



Kiri
Wagstaff

Associate Program Chairs

Workflow Chairs



Zichao Hu



Erich Merrill



Anthony
Opipari



Adam
Parker



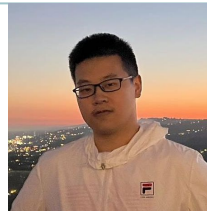
Shelia
Schoepp



Gautham Vasan



Arthur Zhang



Haoran Zhang

AAAI-26 Author Growth



Number of authors in 2026	77302
Number of authors in 2025	46050
Number of authors common to both years	19746
Number of authors only in 2026	57556
Number of authors only in 2025	26304

workflowchairs26@
aaai.zendesk.com



Marc
Pujol- Gonzalez



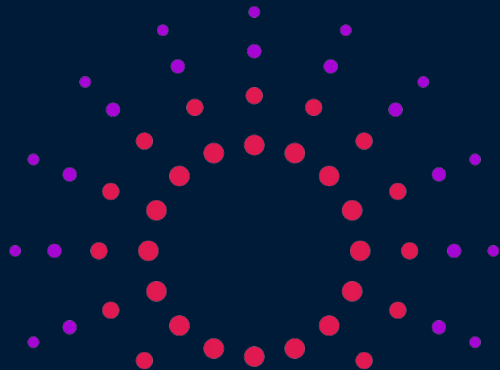
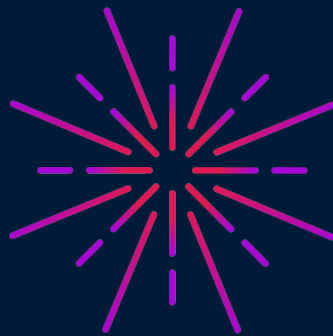
Session Overview

AAAI-26 Main Track Review Process Overview

AAAI-26 AI Reviewing Pilot Program and Findings

Panelist Questions and Insights

Open Question and Answers



Thank You!

AAAI-26 Sponsors – including OpenAI



DIAMOND



PLATINUM



GOLD



SILVER



BRONZE



AAAI-26

Main Track

Review Process



AAAI-26 Review Process



Paper Submissions through OpenReview

Program Committee Formation – PCs, SPCs, and ACs

Program Committee Bidding for Review Assignments

Multi-factor Matching of Papers to Program Committee

Two Phase Reviewing – including Author Response & Discussion

Compliance – including Publications Ethics & Dual Submissions

Panic! Breathe. Adapt.



AAAI-25 had 15,532 submissions – 11,263 were “active submissions”
Expanded Program Committee needed for AAAI-26

Program Committee Recruitment:

accepted / invited

7782 / 22432

Senior Program Committee Recruitment:

accepted / invited

869 / 2773

Area Chair Recruitment:

accepted / invited

267 / 524

Active Submissions:

28713

Withdrawn Submissions:

0

Desk Rejected Submissions:

0

Screenshot from OpenReview panel from Aug 1, 2025

AAAI-26 Program Committee



26,421 Program Committee Members

- Paper review and contribute to summary discussion
- ~11,000 volunteers and ~15,000 reciprocal reviewers

1,728 Senior Program Committee Members

- Review summary and moderation of reviewer discussion
- Decision recommendation

471 Area Chairs

- Broader view over area of submissions and exception handling
- Decision recommendation

2-Phase Review Process



“Phase 0” — summary rejection for invalid submissions

Phase 1 — downselect for submissions unlikely for acceptance

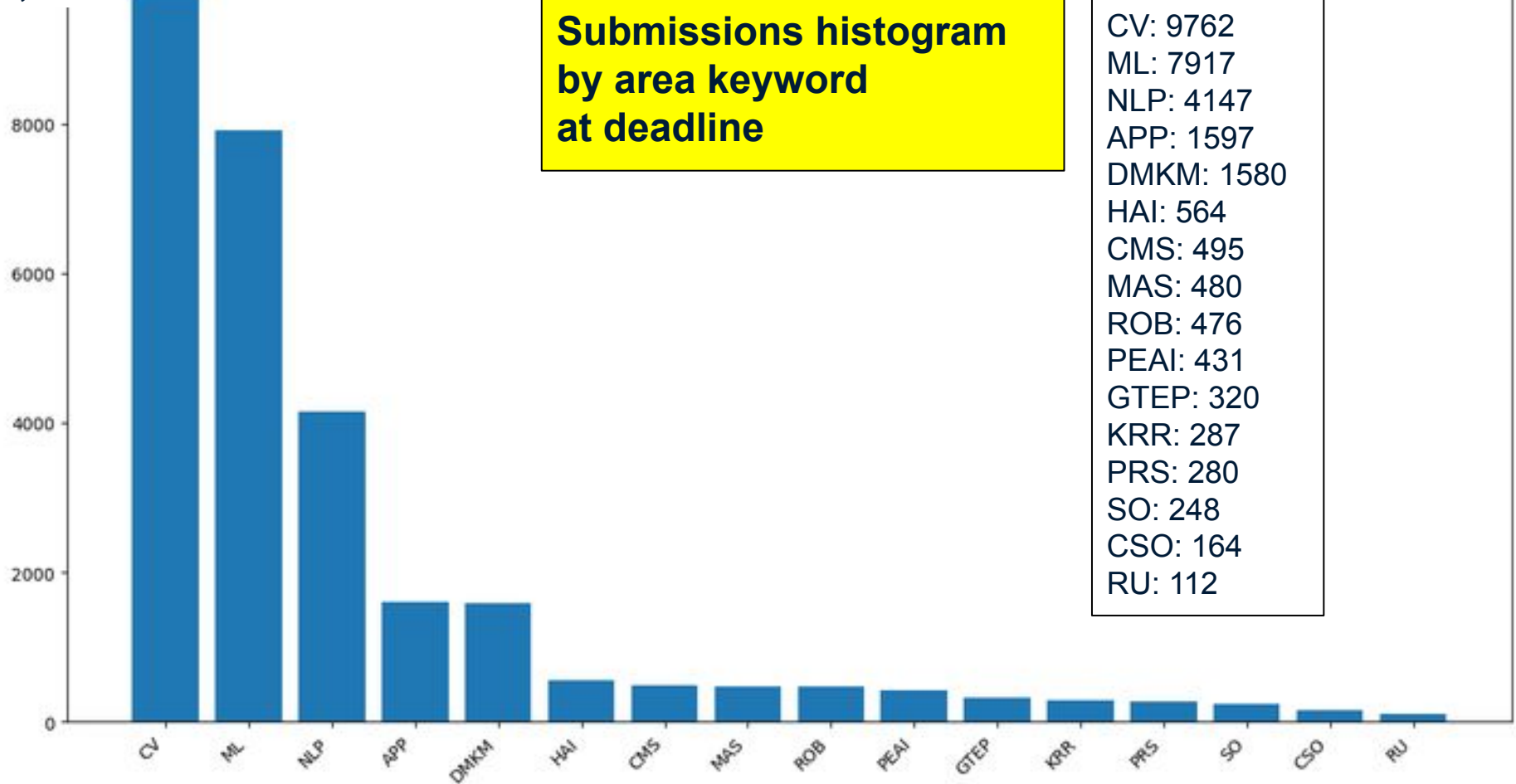
- 2 human reviews (at minimum, 1 reciprocal review maximum)
- 1 AI review
- Downselect recommendations made by SPC and AC
- Promotion to Phase 2 in cases of uncertainty

Phase 2 — final paper decisions

- At least 3 human reviews (most papers had 4 or more reviews)
- Author rebuttal, reviewer discussion with AI summary, and review updates
- SPC, AC makes recommendations

Top 20 Major Keywords in AAAI 2026 Submissions

10,000

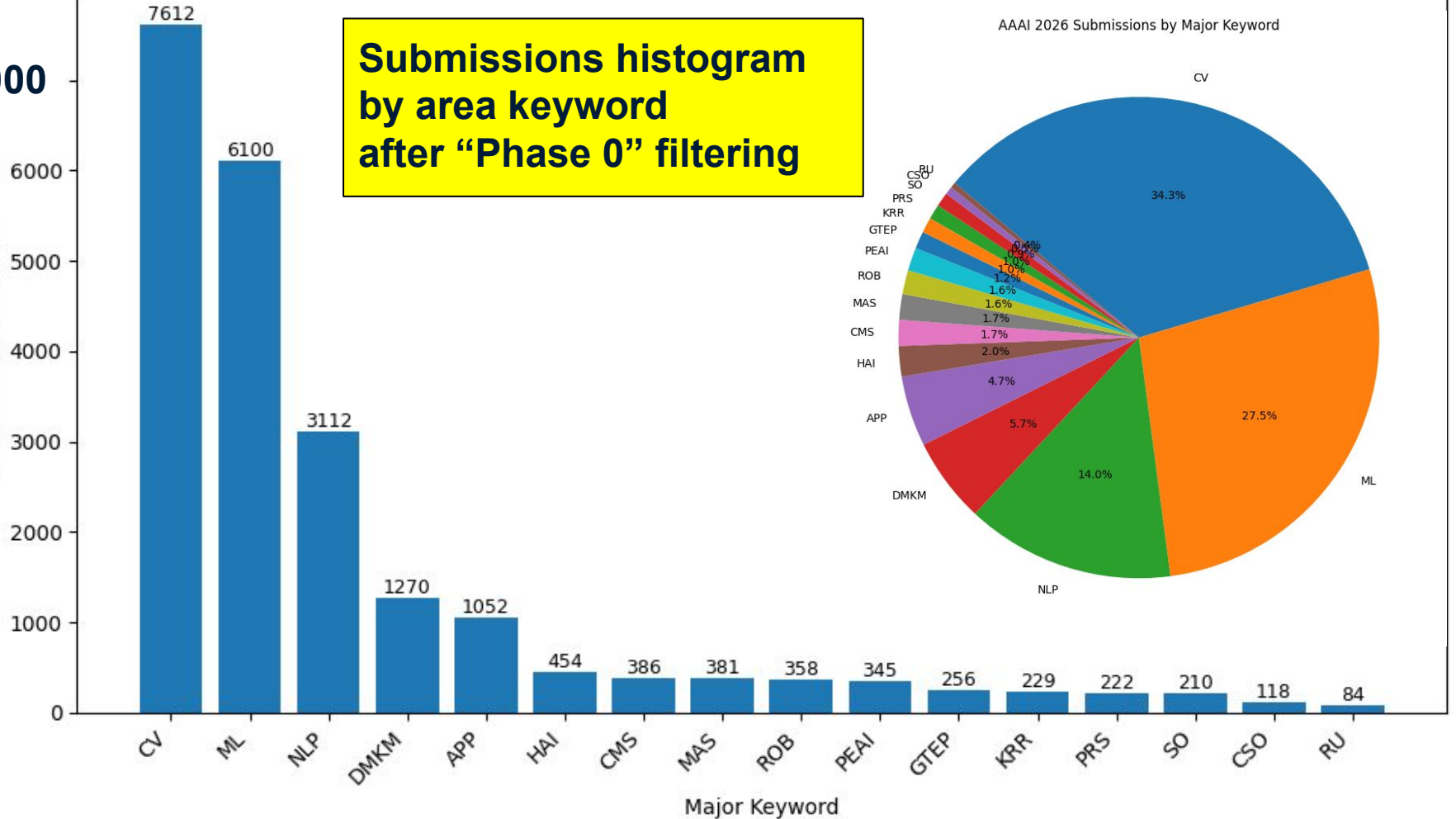


AAAI 2026 Active Submissions by Major Keyword

7,000

**Submissions histogram
by area keyword
after “Phase 0” filtering**

Number of Submissions



Paper-Reviewer Matching



Improving Robustness in Paper-Reviewer Assignment for AAAI 2026

Overview

Paper-reviewer assignment sits at the heart of the conference peer-review process. When done well, it ensures that submissions are evaluated by reviewers who are both qualified and willing to engage.

As conferences continue to grow in size, however, assignment algorithms face new challenges. Scale has increased dramatically, and with it the risk of strategic or coordinated behavior. Modern assignment methods must therefore do more than simply maximize reviewer-paper similarity: they must also be robust to manipulation, promote diversity among reviewers, and at the same time remain computationally efficient.

For AAAI 2026, we designed a new paper-reviewer assignment algorithm with these goals in mind.

Problem Setting and Default Algorithm Used in AAAI 2025

In the standard setting,

- There is a set of **papers** P with n_p papers and a set of **reviewers** R with n_r reviewers.
 - Each paper p must be assigned exactly ℓ_p reviewers.
 - Each reviewer r can review at most ℓ_r papers.
- Assignments are represented by a **binary matrix** $x \in \{0, 1\}^{n_p \times n_r}$, where:
 - $x_{p,r} = 1$ means reviewer r is assigned to paper p .



Fei
Fang

<https://hackmd.io/cY773D0KRwK6XTU0pktVnQ>

Multi-factor Paper Matching

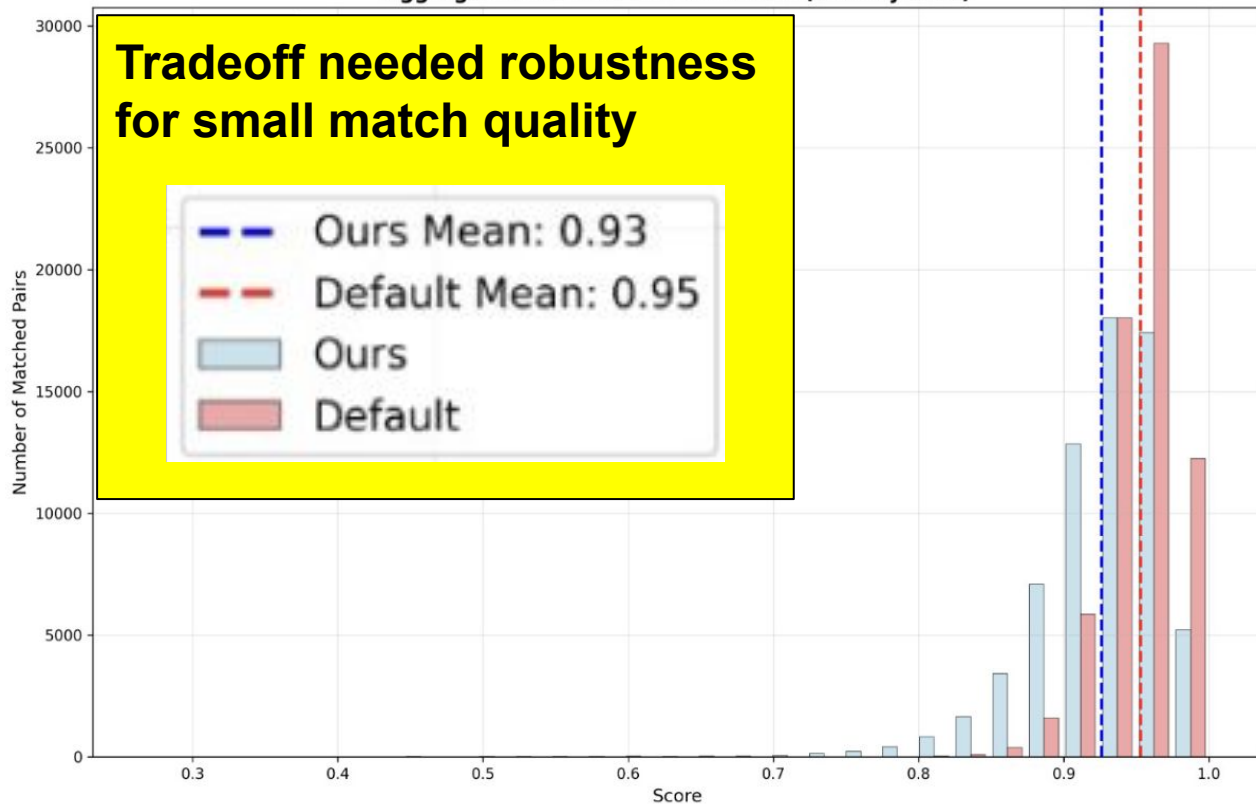


Property	Description	Notes
Quality	Ensure high matching quality measured by the total similarity score	
Efficiency	The matching algorithm should complete within a reasonable time.	
COI	Reviewers must not be assigned to papers with COI.	Hard constraint
NumReviewers	Each paper must be assigned to at least a specified number of reviewers.	Hard constraint
LoadLimit	No reviewer should be assigned more than a specified number of papers.	Hard constraint
Seniority	Each paper must have at least one experienced reviewer.	Important for quality control.
CoauthorDist	No two reviewers assigned to the same paper have co-authored a paper before or have a small distance in the co-authorship graph.	Ensure diversity of views and mitigate collusion
GeoDiversity	Reviewer of the same paper should ideally come from different geographic regions.	Supplementary diversity goal.
2Cycles	Pairs of reviewers who bid positively on each other's papers should not be assigned to review those papers.	Additional safeguard to mitigate potential bid collusion.
Randomized	Randomized algorithms provide more robustness against malicious behavior, and improve reviewer diversity and reviewer anonymity.	

Paper Matching Results

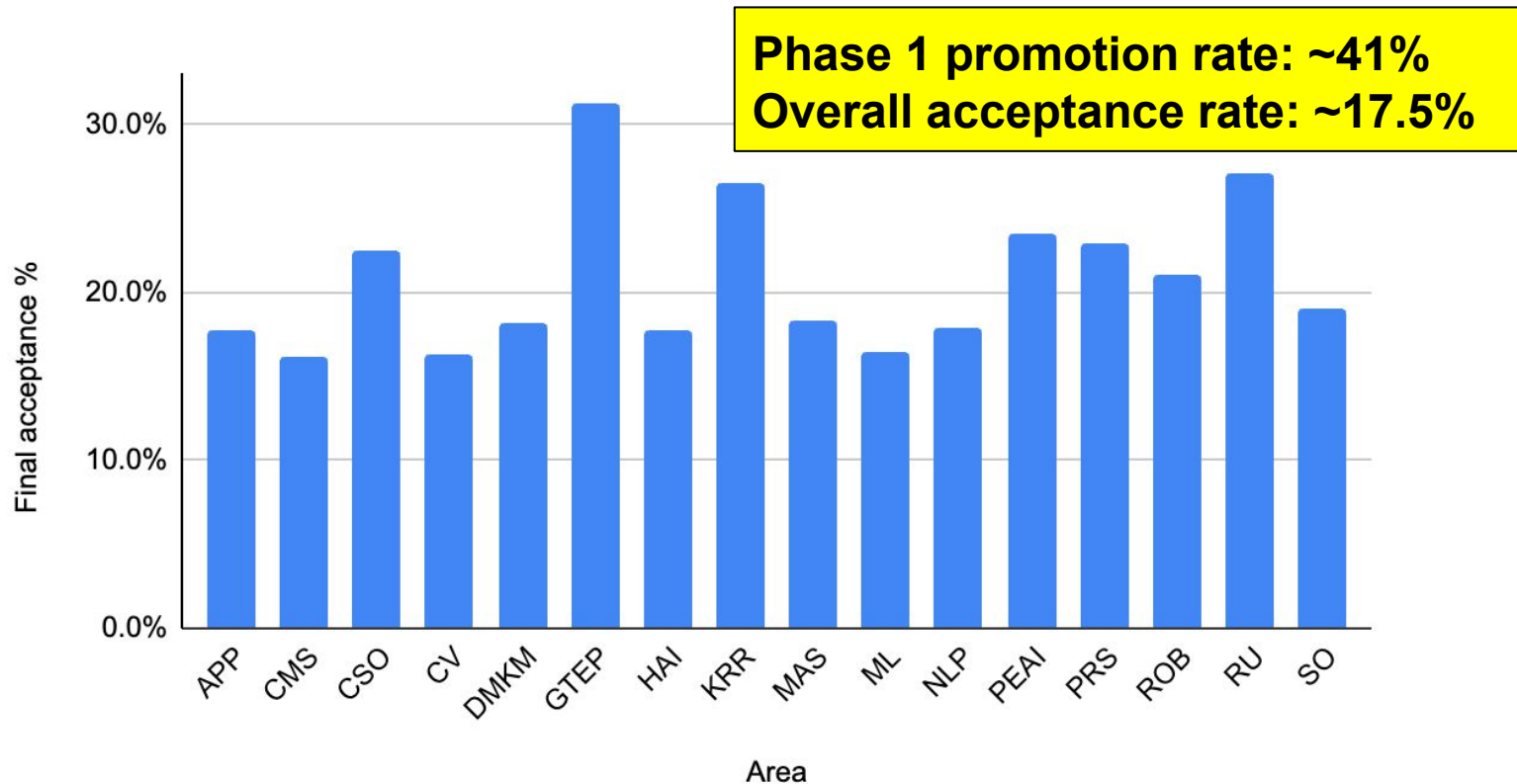


Aggregate Scores of Matched Pairs (Side-by-Side)



- Greater robustness to bidding manipulation
- Better preservation of COI and load constraints
- Give up a small bit of matching quality score

Main Track Acceptance Rate



Dual Submissions Compliance

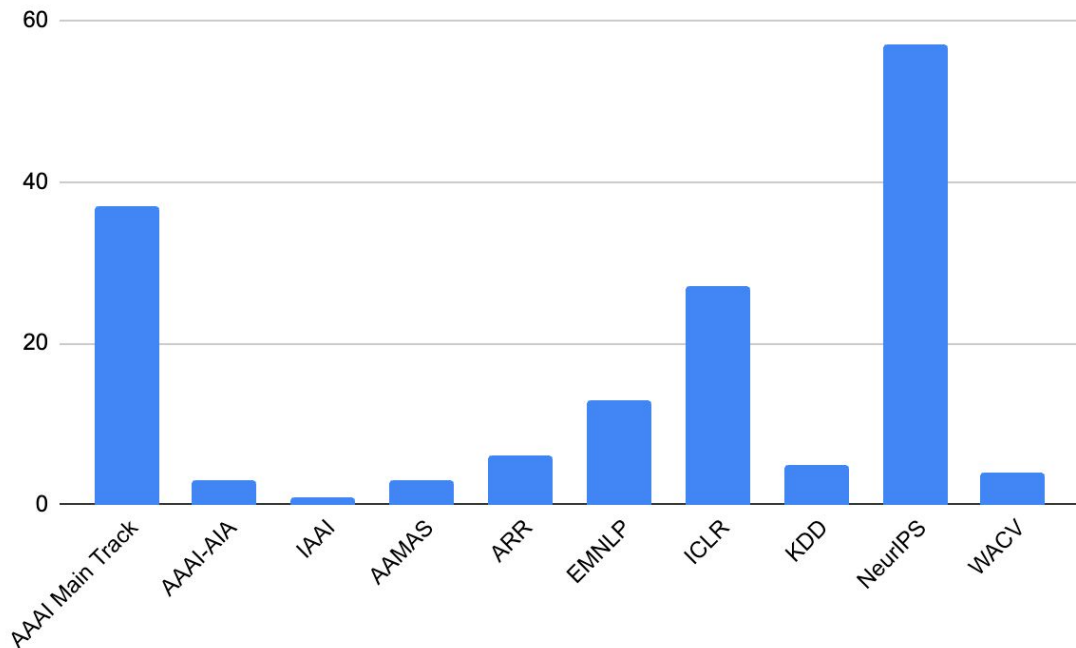


156 dual submissions rejected

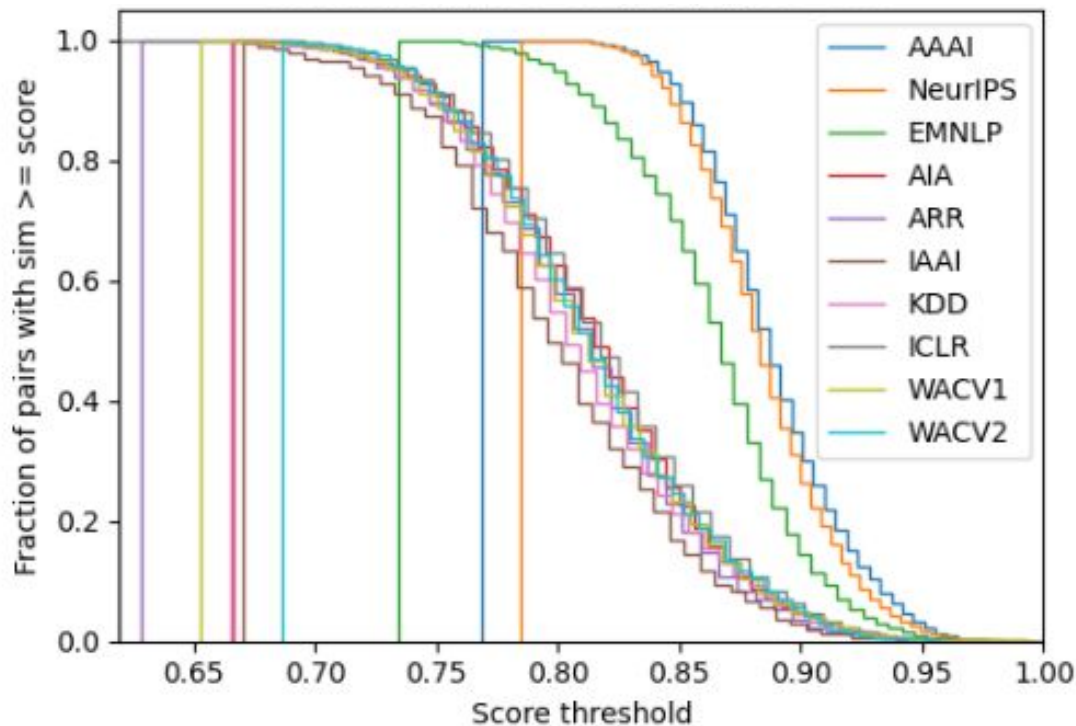
- Abstract & title similarity
- LLM check
- **3 humans** verify



Kiri
Wagstaff



Collaboration between conferences for dual submissions compliance





Matt Taylor Jun 23rd, 2025 at 12:17 PM

image.png ▼



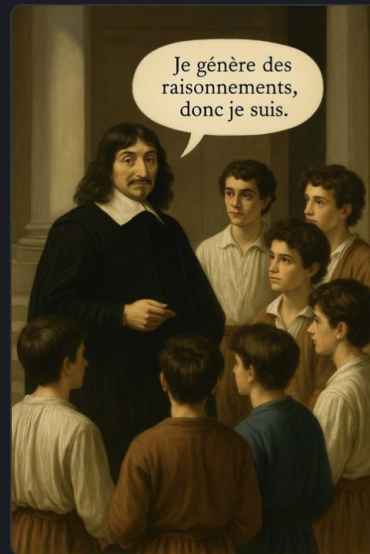
3 replies



Joydeep Biswas Jun 23rd, 2025 at 2:38 PM

I shared it with my friend who I co-authored the LLM reasoning eval paper with, and he pointed out that it's the wrong millenium

image.png ▼



Chad Jenkins Jun 23rd, 2025 at 3:00 PM

I guess Matt put Descartes before the horse



Joydeep Biswas Jun 23rd, 2025 at 3:17 PM

And Chad's back in the game! :D (edited)

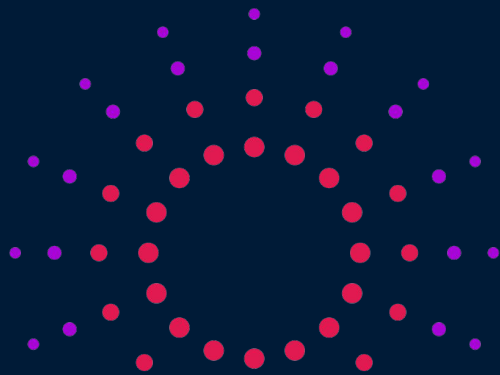
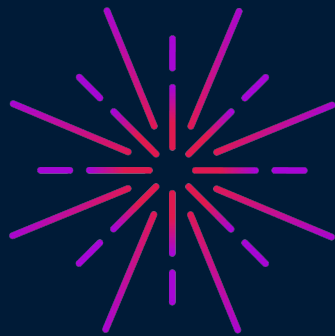


Q&A – Post your questions to Slido



AAAI-26

AI Reviewing Pilot Program



Open Questions



1. Can state-of-the-art AI systems generate high-quality peer reviews of scientific papers?
2. How do such state-of-the-art AI-generated reviews compare to human-written reviews?
3. How can AI tools be effectively integrated into the review process?

AAAI-26 AI Review Pilot Program



Key question: Can AI tools assist in the scientific peer review process?

Process:

- One **additional, clearly labelled** AI-generated review in phase 1
- No **scores** or **recommendations** by AI review
- No human replaced in the review process
- Reviewers (after submitting their own) SPCs, and ACs could view the AI review

Technical Details:

- AI review generated by **custom, multi-stage** process
- Process included structured conversion including math & tables
- Tool use included code interpreter, literature search
- Base LLM: OpenAI GPT5, with **Zero Data Retention**

What We Built



AAAI-26 AI Review Generation: A Multi-Stage Process



- Paper pre-processing:
 - Resampled to convert images (if present) to a fixed and consistent resolution
 - Converted into structured markdown using a machine learning based OCR to extract math in latex notation, and table structures
- A preliminary analysis of the story of the paper
- Technical accuracy checking, code interpreter available as tool
- Literature search using generated keywords, queries, and a web search tool
- Results checking, including all figures, tables, and text descriptions
- Self-critique: Check the review itself to check for factual accuracy, consistency with the paper, and any unsubstantiated claims. Revisions are made to address issues found.

AI Review Generation Dashboard

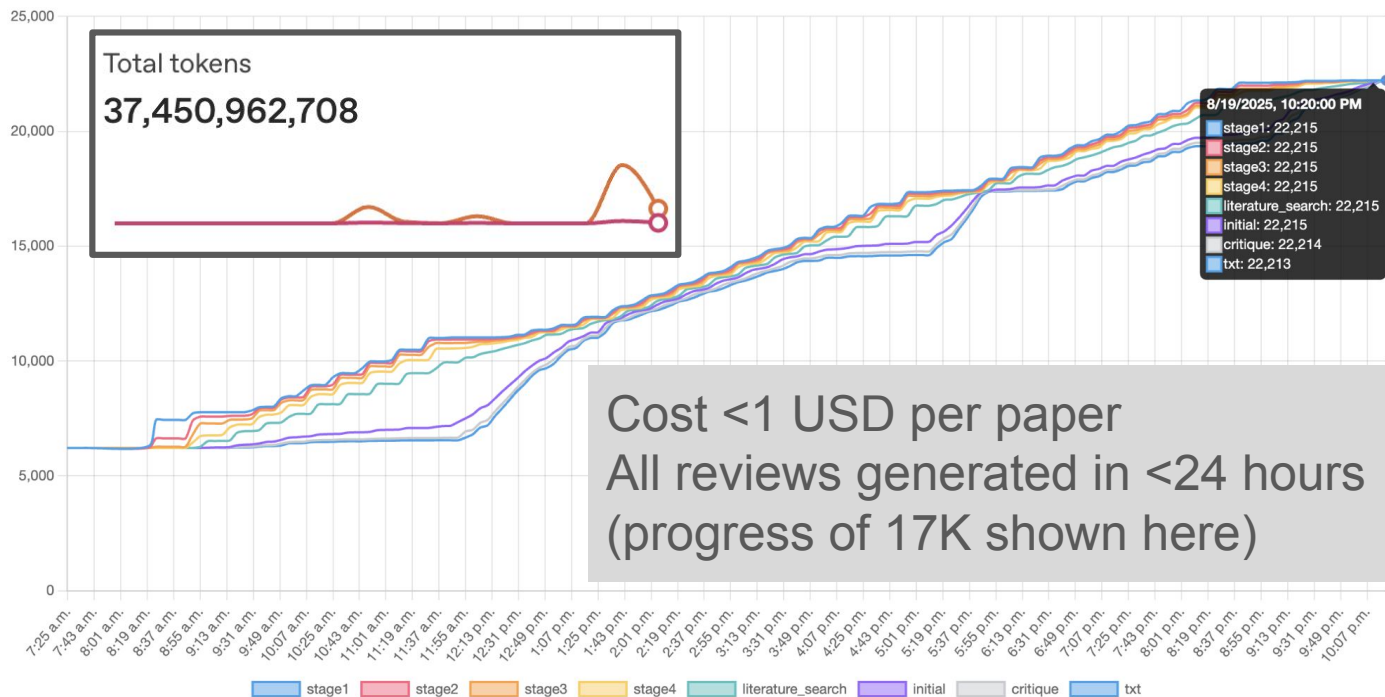


AAAI 2026 AI Review Status Dashboard

Last update: 8/19/2025, 10:25:00 PM (auto-refreshing)

Time window:

Scroll/pinch to zoom, Shift-drag to box-zoom, Double-click to reset zoom



Unique Challenges Discovered



Challenges:

- Token limits for full papers
- Errors in reading equations and tables
- Discrepancy between OCR and PDF
- Cost and time in scaling

Time per paper (H:M:S)	0:25:00			
Number of papers	22217			
RPM Limit (Tier 5)	15,000			
TPM Limit (Tier 5)	40,000,000			
Parallelism	128	256	512	1024
Total time	72:19:15	36:09:38	18:04:49	9:02:24
RPM/request/minute	117	59	29	15
TPM/request/minute	312,500	156,250	78,125	39,063

Empirically observed to be less of a concern:

- Blatant hallucinations
- Making up citations
- Revealing author identities

Details



- OCR Model: olmOCR, <https://olmocr.allenai.org/>
- LLM Model: GPT-5, Zero Data Retention, High reasoning effort, Flex tier
- Tools: code interpreter, web search
- 512x parallelism with interruptible workflow, checkpointing, graceful backoff and retry
- Each paper review takes ~20 minutes to complete
- OCR took ~12 hours on one 8xH100 server
- Review generation: <24 hours

Oversight



- Automated quality checking: a critic LLM checks for specific concerns
- Specific checks for intrusion checking
- Manual Oversight:
 - We manually sampled reviews at every stage of deployment
 - We requested SPCs to identify issues - and they successfully identified

Quality Checking



Policy violations:

- q1: Does this review reveal the author identity?
- q2: Does this review raise ethical concerns?
- q3: Does the review raise any policy violations (e.g. double-blind policy violations)?

Appropriate:

- q4: Does this review contain potentially offensive content?
- q5: Does the review make unfair judgments based on affiliation, geography, gender, or other irrelevant factors?
- q6: Is the review excessively subjective without justification (e.g., "I don't like this area of research")?

Quality:

- q7: Should this review be flagged for being incomprehensible or otherwise poorly written?
- q8: Should this review be flagged for being technically incorrect?
- q9: Does this review include clearly labelled sections: "Title", "Synopsis of the paper", "Summary of Review", "Strengths", "Weaknesses", "Suggestions for Improvement", "References"?
- q10: How would you rate the overall quality of this review? (Poor / Fair / Good / Excellent)

Misc:

- q11: Should this review be flagged as problematic for any reason?
- q12: Does this review appear to be generated by an unqualified reviewer?
- q13: Does this review appear to be generated by an LLM?
- q14: On a scale of 0 to 10, how much effort do you think the reviewer put into this review? E.g. 0 = a review by an unqualified reviewer, 5 = a typical average AAAI reviewer, 10 = an outstanding reviewer award candidate.

Energy Use: Coarse Approximation



Dominant Compute Usage:

- Conversion to markdown: GPU
- Multi-stage review process: GPU

Assumptions For Estimation:

- LLM Model size: 4T parameters, 8-bit quantization
- GPU, Usage: H100, 100% usage (8x8xH100 rack per model)

Total CO2 emissions:

- 148,939.95 pounds CO2
- Miles equivalent driven in a 30mpg gasoline car: 229,138
- Number of round-trip passengers from JFK to SIN equivalent: 28

AAAI-26 AI Review Contents



Every AI review includes:

- Brief synopsis of the paper
- Summary of review - in a structured format
- Clear strengths and weaknesses
- Constructive suggestions for improvement
- **Highly detailed feedback** with specific evidence for every point in the above
- Detailed math notation when discussing mathematical and algorithmic specifics

Example Review



Title: Every Bit Helps: Achieving the Optimal Distortion with a Few Queries

Synopsis of the paper The paper studies how a small, fixed number of cardinal utility queries per agent can dramatically reduce worst-case social-welfare distortion in one-sided matching and single-winner voting. It proposes a simple, deterministic, polynomial-time algorithm that, with λ queries per agent, achieves distortion proportional to λ raised to the one over λ in one-sided matching, which is optimal up to constant factors for fixed λ . Using a similar template and stable committees, it achieves distortion proportional to λ raised to the one over λ in one-sided matching, which is optimal up to constant factors for fixed λ . Using a similar template and stable committees, it achieves distortion proportional to λ raised to the one over λ in one-sided matching, which is optimal up to constant factors for fixed λ . The minimum of the number of agents and candidates raised to the one over λ in voting, again optimal up to constants for a fixed number of queries. A key technical ingredient is a k -capacity serial dictatorship that constructs an exactly stable set of matchings (and approximately stable committees in voting) to guide which values to query.

Summary of Review This is a strong and timely paper that cleanly generalizes the recent “two queries suffice” results to any fixed number of queries, with tight bounds (up to constants) in both matching and voting. The algorithmic ideas are simple yet powerful, and the technical contribution establishing the existence of exactly stable sets of matchings is elegant and potentially of independent interest. The proofs for matchings are clear; for voting, the main algorithm and most of the proof are deferred to the full version, which somewhat limits verifiability from the conference submission alone. Some presentation issues (minor typos, rounding, and notation) can be addressed in a revision. Overall, the work appears sound, advances the state of the art, and should be of broad interest to the community.

Strengths

- Technically significant bounds with few queries:
 - One-sided matching: a deterministic, polynomial-time mechanism using λ queries per agent achieves distortion $\lambda \cdot n^{1/\lambda}$ (Theorem 5). This strictly generalizes the $\Theta(\sqrt{n})$ two-query guarantee of Amanatidis et al. (2024) and matches the $\Omega(n^{1/\lambda})$ lower bound of Amanatidis et al. (2022) up to constants for constant λ .
 - Single-winner voting: a deterministic mechanism achieves $O(\lambda \cdot (\min n, m)^{1/\lambda})$ distortion (Theorem 6), optimal up to constants for fixed λ given the $\Omega(m^{1/\lambda})$ lower bound in Amanatidis et al. (2024).
- Clear algorithmic template with principled query design:
 - The paper introduces a k -capacity serial dictatorship subroutine (Algorithm 1) that, together with a decomposition lemma, yields an exactly stable set of k matchings for any $k \in [n]$ and positive agent weights (Theorem 4). This is conceptually powerful and appears novel in the matching context (in contrast to only approximately stable committees in multiwinner voting (Jiang, Munagala, and Wang, 2020)).
 - The λ -round mechanism (Algorithm 2) uses the queried values as weights to build a geometrically decreasing sequence of stable objects (matchings or committees), and then solves a final welfare maximization with conservative underestimates u^* of the utilities. The analysis by grouping agents by the first round at which they would “deviate” is crisp and modular.
- Optimality and improved tradeoffs:
 - For matching, the paper not only settles the optimal $\Theta(n^{1/\lambda})$ distortion (up to constants) for constant λ , but also obtains $O(\log n)$ distortion with $\log n$ queries deterministically (improving the query complexity over Amanatidis et al. (2021) for this target).
 - For voting, the dependence on $\min n, m$ rather than n strengthens the guarantees in regimes with many agents and few candidates and supports compelling reductions (e.g., viewing matchings as “candidates”) with only constant factor blow-up in distortion.
- Clarity of key ideas and connections:
 - The paper effectively unifies and extends techniques from matching and multiwinner voting (e.g., stable committees (Cheng et al., 2020; Jiang, Munagala, and Wang, 2020; Aziz et al., 2017)) to derive tight query-distortion tradeoffs without normalization or randomization assumptions.

Weaknesses

- Missing core details for voting in the main text: The algorithm and most of the proof for Theorem 6 are deferred to the full version. Given the centrality of this result, including at least pseudocode and a more detailed proof sketch in the conference submission would materially aid verification and reproducibility.
- Large hidden constants via approximate stability: The voting analysis relies on β -approximately stable committees from Jiang, Munagala, and Wang (2020), with β around $32 + \epsilon$ (or 16 for identical weights). While this does not affect asymptotics, the constants may be sizable in practice and deserve explicit discussion.
- Minor notational and presentation issues:
 - In Algorithm 2 and Lemma 2, there is a likely notational slip: “ $g_1(i) = \sigma_1(i)$ ” should read $g_1(i) = \sigma_1(i)$ (the top choice of agent i). Similarly, in the description of the first round, “since $\lambda = 1$ ” seems to mean “for round $\ell = 1$ ”.
 - Committee size and capacity parameters such as $k \leftarrow n^{1-(\ell-1)/\lambda}$ (or $(\min n, m)^{1-(\ell-1)/\lambda}$ in voting) should explicitly specify rounding and argue that rounding only affects constants.
 - The stability proof in Theorem 4 uses strict inequalities $w(i') > w(i)$ when discussing earlier agents; equal-weight tie-breaking should be addressed explicitly (the argument still goes through with $\geq$ and a positive-weight slack, but stating this would remove ambiguity).
- A potentially confusing statement: Following Theorem 6, the paper writes “More precisely ... the distortion guarantee is $\beta \cdot 1/\lambda \cdot (\min n, m)^{1/\lambda}$.” This appears at odds with the headline $O(\lambda \cdot (\min n, m)^{1/\lambda})$ bound and with the matching analysis, where the grouping argument introduces a factor linear in the number of rounds. If the $1/\lambda$ is intended (perhaps due to a refined averaging), it merits a short derivation; otherwise it may be a typo that should be corrected.
- Scope and assumptions: The analysis assumes strictly positive utilities (used in Theorem 4 via $u_{\min} > 0$). It would help to spell out whether allowing zero utilities affects any steps, and whether small relaxations (e.g., non-negativity) preserve the guarantees.

Suggestions for Improvement

- Strengthen the single-winner section in the main paper:
 - Include pseudocode for the voting analogue of Algorithm 2 and a clear definition of the per-round committee construction, weights, and queries. Add a concise proof sketch identifying the key invariants (committee stability guarantee used per group, geometric schedule, and the final averaging argument), so the result is verifiable without relying on the external full version.
 - Clarify the “more precisely” bound after Theorem 6: if the factor is indeed β/λ , explain the mechanism by which the factor of λ cancels; if not, correct this to $\beta \cdot \lambda$ and note the constant factor explicitly.
- Make the matching stability proof airtight:
 - State the tie-breaking policy for equal weights and adjust inequalities to $\geq$ where appropriate; point out that the positive-weight slack restores a strict inequality needed for the final contradiction.
 - Clearly specify that the weights are strictly positive and comment on whether the proof adapts to non-negative weights.
- Notation, rounding, and readability:
 - Specify the rounding for capacities and committee sizes and note that it only affects constants in the distortion. Consider adding a small figure or example that walks through two or three rounds of the algorithm, showing how the queried values propagate to u^* and why this yields a lower bound on true utilities for higher-ranked options.
 - Standardize the notation $a <_i a'$ (used to mean $a = a'$ or agent i prefers a to a') early and remind the reader where it is employed (notably in constructing u^*).
 - Fix minor typos (“since $\lambda = 1$ ” in the first-round discussion, and “ $g_1(i) = \sigma_1(i)$ ”).
- Contextualization and constants:
 - Add a short paragraph explicitly comparing the number of queries needed to reach $O(\log n)$ distortion deterministically here versus prior work (Amanatidis et al., 2021), emphasizing the improvement in query complexity for the same target.
 - Discuss the impact of the stability constant β on practical performance and whether using improved stability approximations (e.g., the tighter factor for identical weights) could systematically lower constants in the proposed voting algorithm.

References

- Amanatidis, G., Birmpas, G., Filos-Ratsikas, A., & Voudouris, A. A. (2021). Peeking behind the ordinal curtain: Improving distortion via cardinal queries. *Artificial Intelligence*, 296, 103488.
- Amanatidis, G., Birmpas, G., Filos-Ratsikas, A., & Voudouris, A. A. (2022). A few queries go a long way: Information-distortion tradeoffs in matching. *Journal of Artificial Intelligence Research*, 74, 227–261.
- Amanatidis, G., Birmpas, G., Filos-Ratsikas, A., & Voudouris, A. A. (2024). Don’t Roll the Dice, Ask Twice: The Two-Query Distortion of Matching Problems and Beyond. *SIAM Journal on Discrete Mathematics*, 38(1), 1007–1029.
- Aziz, H., Elkind, E., Faliszewski, P., Lackner, M., & Skowron, P. (2017). The Condorcet principle for multiwinner elections: from shortlisting to proportionality. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence* (pp. 84–90).
- Caragiannis, I., & Fehr, K. (2023). Beyond the worst case: Distortion in impartial culture electorate. *arXiv preprint arXiv:2307.03750*.
- Cheng, Y., Jiang, Z., Munagala, K., & Wang, K. (2020). Group fairness in committee selection. *ACM Transactions on Economics and Computation*, 8(4), 1–18.
- Ebadian, S., Kahng, A., Peters, D., & Shah, N. (2022). Optimized distortion and proportional fairness in voting. In *Proceedings of the 23rd ACM Conference on Economics and Computation* (pp. 563–600).
- Jiang, Z., Munagala, K., & Wang, K. (2020). Approximately stable committee selection. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing* (pp. 463–472).

Review for AAAI-2025 Best paper award winning paper
<https://ojs.aaai.org/index.php/AAAI/article/view/33507>

Example Metareview



Points of Agreement:

- Strengths: Reviewers X, Y, Z, agree that ...
- Weaknesses: Consensus among all reviewers about...

Points of Disagreement:

- While some reviewers (A, B, C) appreciate the thoroughness of the experiments, reviewer D highlights a potential missing baseline...

Overall, the proposed approach [...] is significant in [...], but could be further strengthened by [...]

Survey Responses And Analyses



Structure Of Survey



- For THIS review:
 - Overall impression
 - Review emphasis
 - Technical accuracy
 - Suggestions for research and writing
- General impressions
 - Overall assessment of AAAI-26 AI reviews
 - Sentiment towards AI in future peer review

Study reviewed by UT Austin, UMich, and UAlberta IRB

Survey Questions



[Overall Impressions]

This **AI-generated** Review:

1. **[For AI review only]** This review demonstrated paper review capabilities beyond what I expected from AI
2. This review was a thorough review for this conference

[Review Emphasis]

3. This review accurately conveyed the significance and impact of the paper
4. This review overemphasized minor issues compared to major contributions
5. This review raised points that I had not previously considered
6. **[For reviewers only]** This review changed my evaluation or interpretation of the paper
7. **[For authors only]****[For AI review only]** This review raised important points that the human-written reviews omitted

[Technical Accuracy]

8. This review accurately identified technical errors in the paper that I had not discovered
9. This review made technical errors in the review or reading of the paper
10. **[For reviewers only]** This review discovered high-level concerns that other human reviewers would have difficulty catching
11. **[For reviewers only]** This review overlooked important points that other human reviewers would likely have caught

[Suggestions For Research and Writing]

12. This review provided useful suggestions to improve the research design/methods.
13. This review provided useful suggestions to improve the presentation of the paper
14. This review provided suggestions that were wrong, infeasible, or otherwise unhelpful
15. This review correctly suggested related work I was previously unaware of
16. **[For AI review only]** This review incorrectly suggested "related work" that is non-existent (i.e., a "hallucinated" citation)
17. **[For authors only]** This review raised points that I want to address in a revision (camera-ready or resubmission)

Survey Response Rates



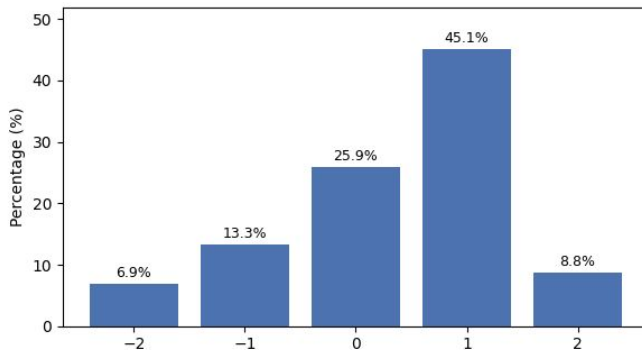
Category	Authors	PC	SPC	AC
AI Reviews	575	1305	117	12
Human Reviews	2500	879	433	13
Total	3075	2184	550	25

Total Survey Responses: 5834

Overall Impressions



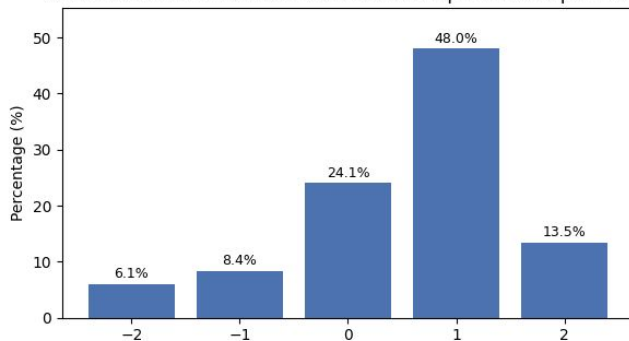
Overall AAAI26 AI Reviews Were Useful



Agree	53.9%
Neutral	25.9%
Disagree	20.2%

Overall AI Reviews Would be Useful in Future Peer Review Process

Overall AI reviews would be useful in future peer review processes



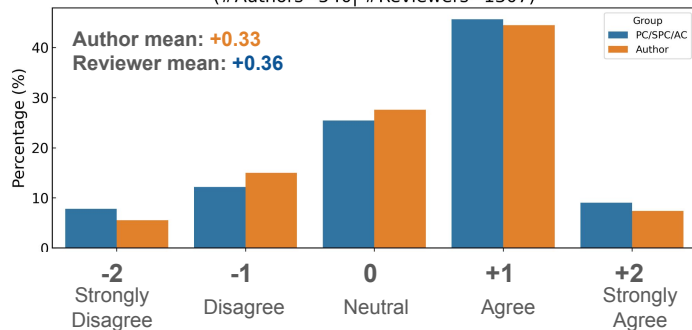
Agree	61.5%
Neutral	24.1%
Disagree	14.5%

Useful vs. Harmful



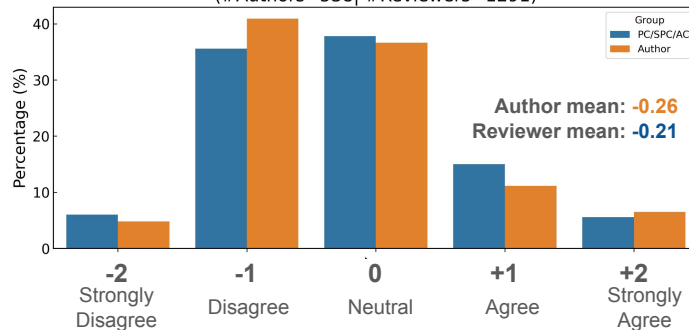
Overall AAAI26 AI Reviews Were Useful

(#Authors=540 | #Reviewers=1307)



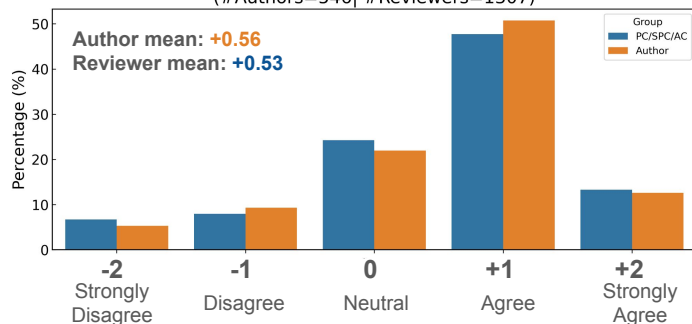
Overall AAAI26 AI Reviews Were Harmful

(#Authors=538 | #Reviewers=1291)



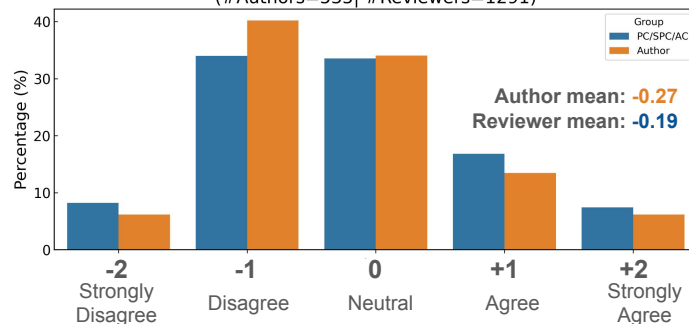
Overall AI Reviews Would be Useful in Future Peer Review Process

(#Authors=546 | #Reviewers=1307)

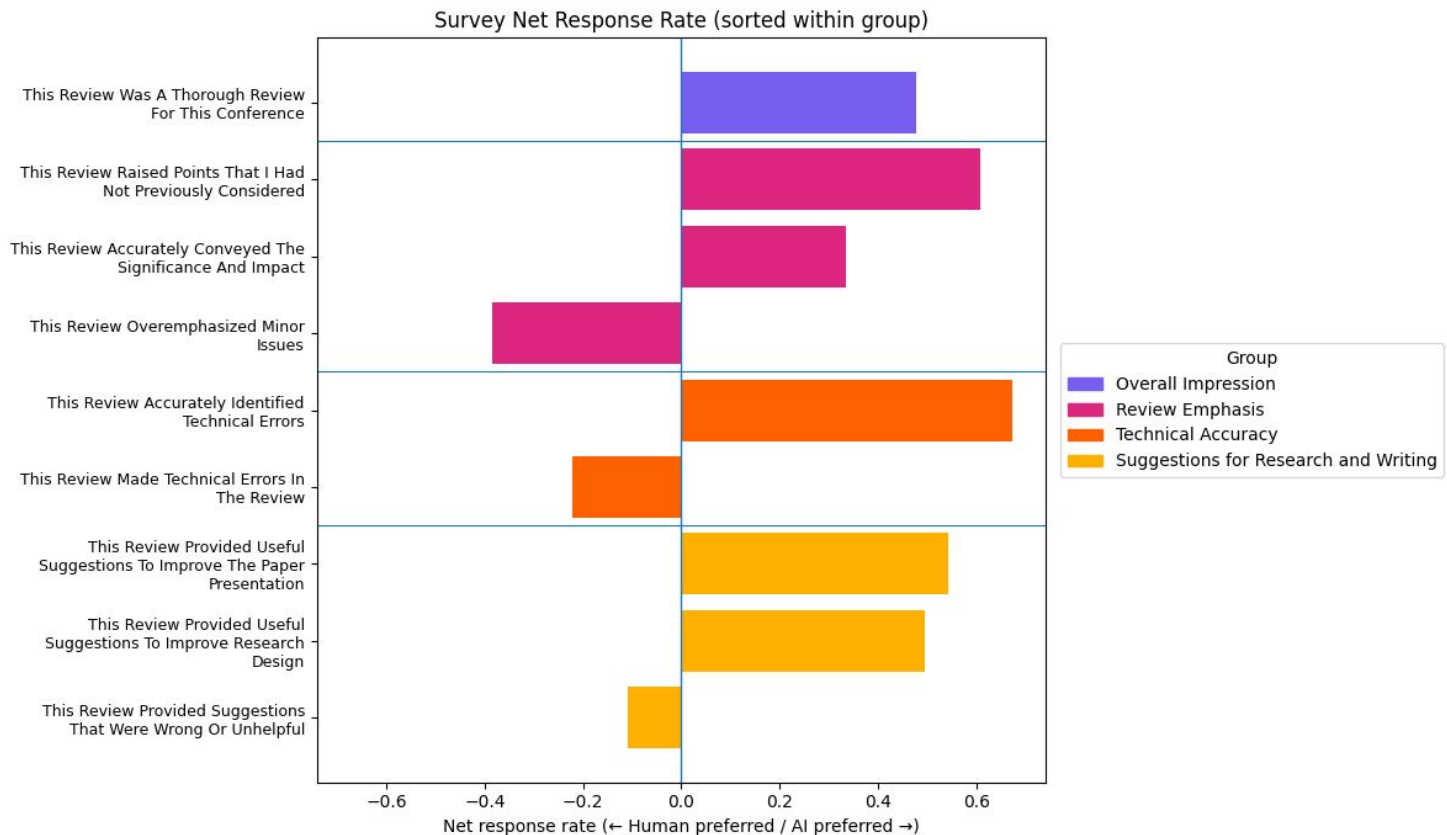


Overall AI Reviews Would be Harmful in Future Peer Review Process

(#Authors=535 | #Reviewers=1291)



Human Vs. AI Reviews - All Responses



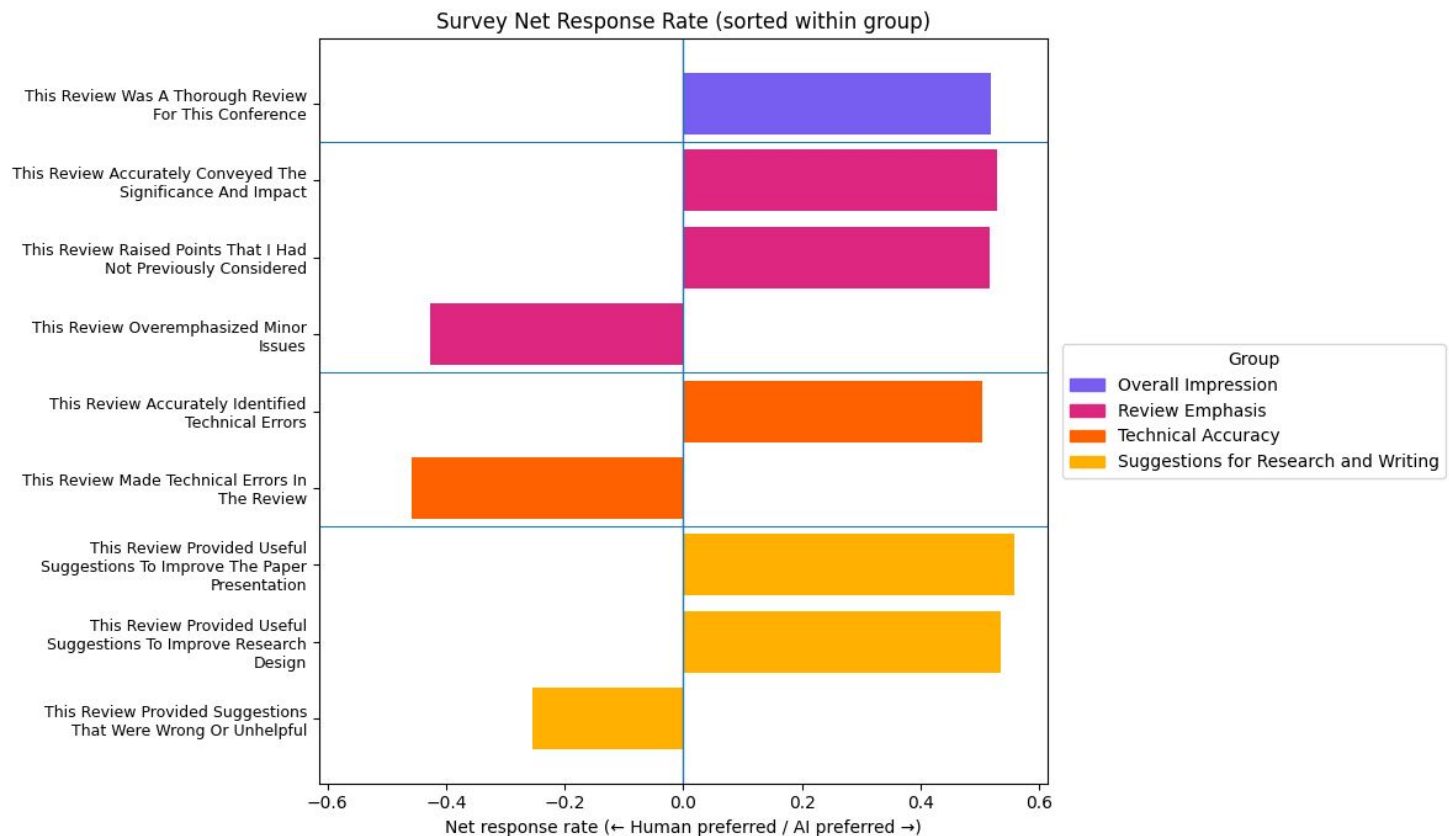
Human Vs. AI Reviews: P-values



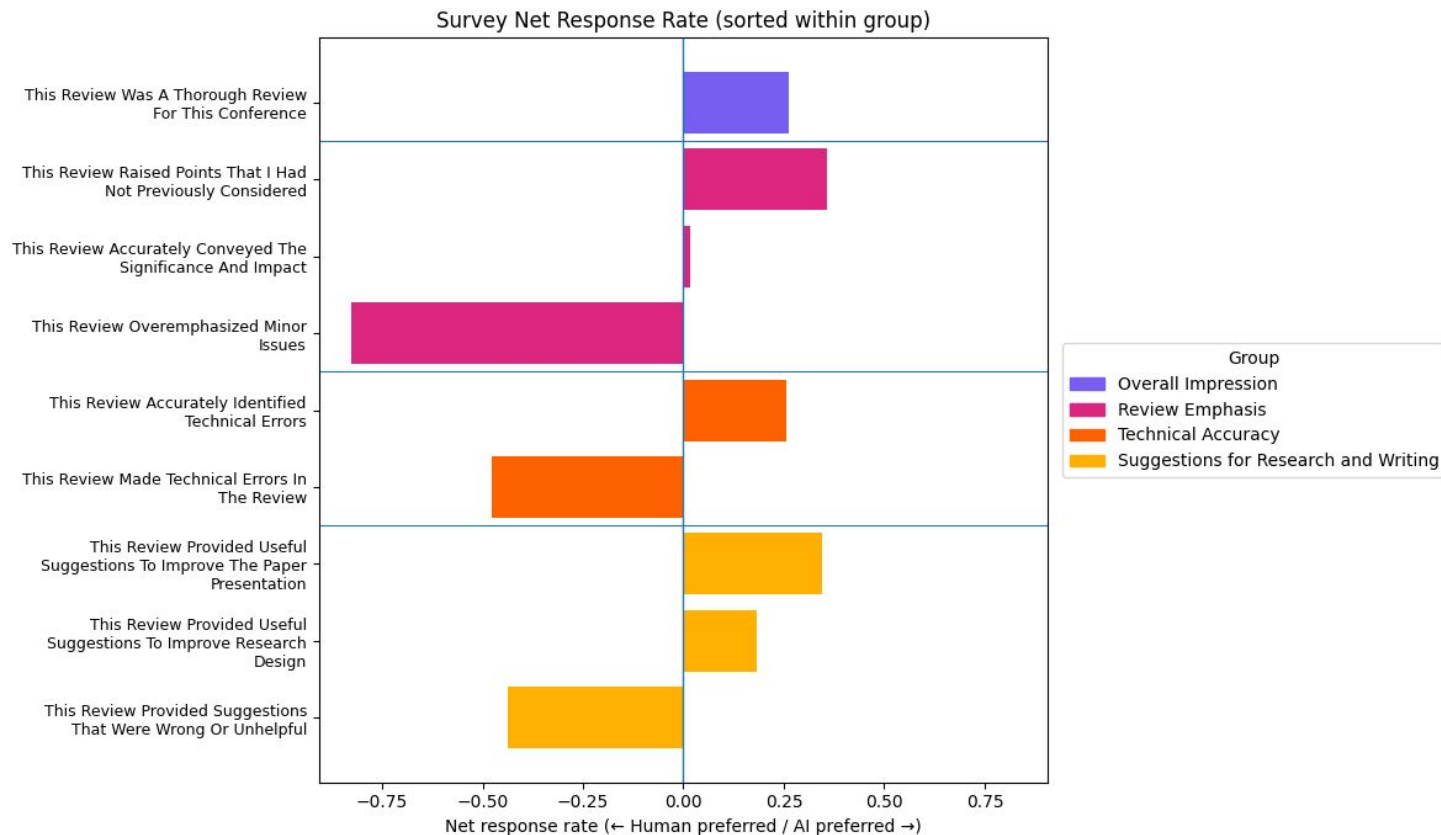
This review provided suggestions that were wrong or unhelpful	5.03E-06
This review provided useful suggestions to improve research design	1.26E-45
This review accurately identified technical errors	6.01E-97
This review made technical errors in the review	1.09E-16
This review overemphasized minor issues	1.83E-30
This review accurately conveyed the significance and impact	2.12E-15
This review raised points that I had not previously considered	2.77E-77
This review provided useful suggestions to improve the paper presentation	3.97E-59
This review correctly suggested related work	1.58E-91
This review was a thorough review for this conference	1.73E-35

Human Vs. AI Reviews - Mann-Whitney U Test P-values

Human Vs. AI Reviews - Author Responses



Human Vs. AI Reviews - PC/SPC/AC Responses

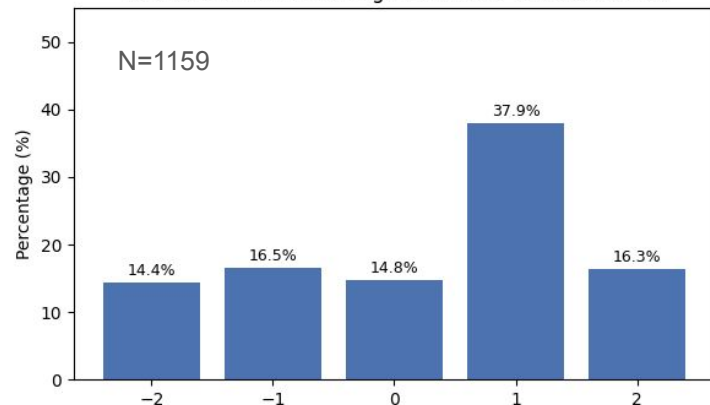


Accept

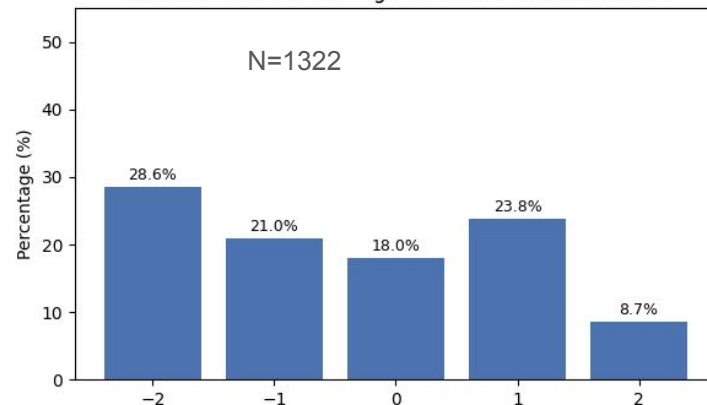
Reject

Human Reviews

Accepted Papers, Author Response: Human-Review
This review was a thorough review for this conference

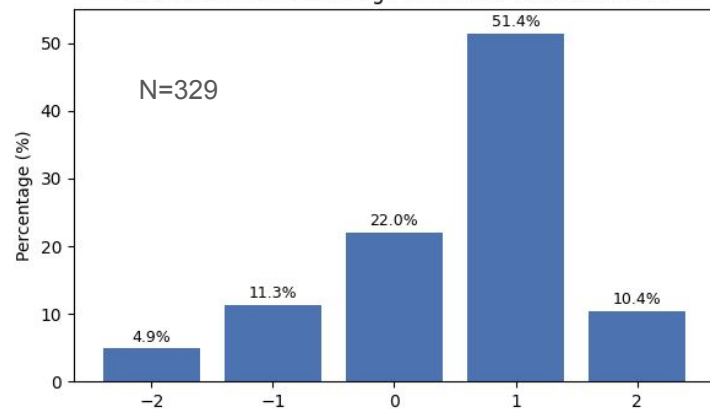


Rejected Papers, Author Response: Human-Review
This review was a thorough review for this conference

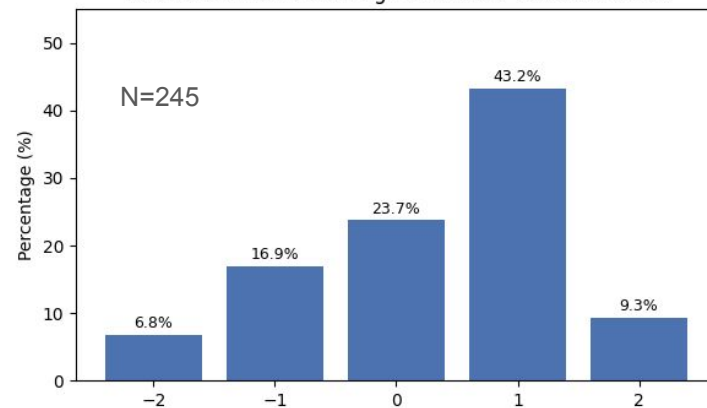


AI Reviews

Accepted Papers, Author Response: AI-Review
This review was a thorough review for this conference

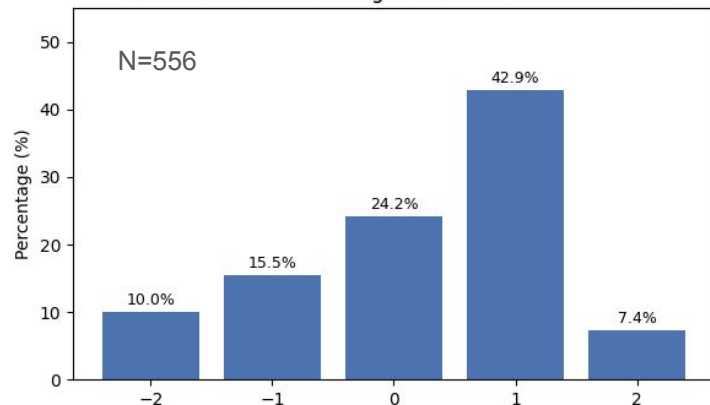


Rejected Papers, Author Response: AI-Review
This review was a thorough review for this conference



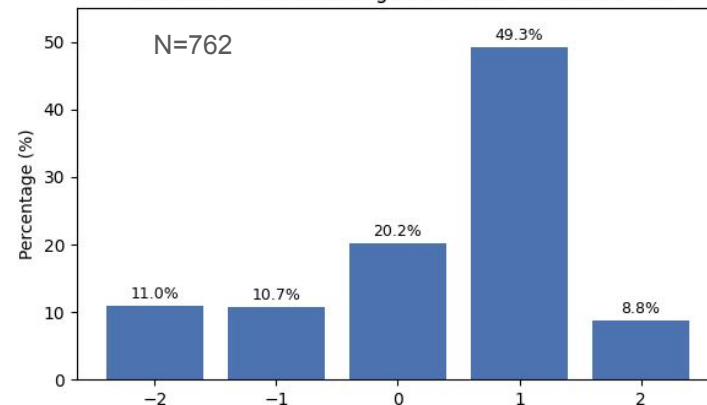
Accept

Accepted papers, PC+SPC+AC: Human-Review
This review was a thorough review for this conference

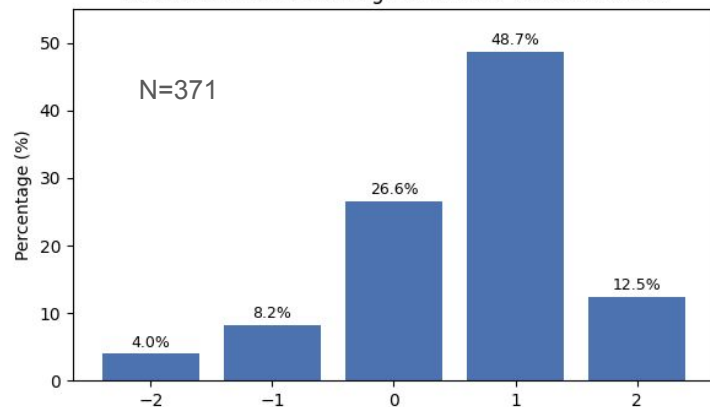


Reject

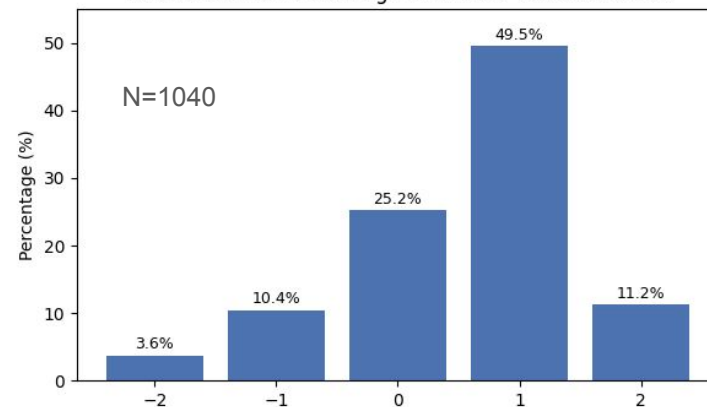
Rejected papers, PC+SPC+AC: Human-Review
This review was a thorough review for this conference



Accepted papers, PC+SPC+AC: AI-Review
This review was a thorough review for this conference



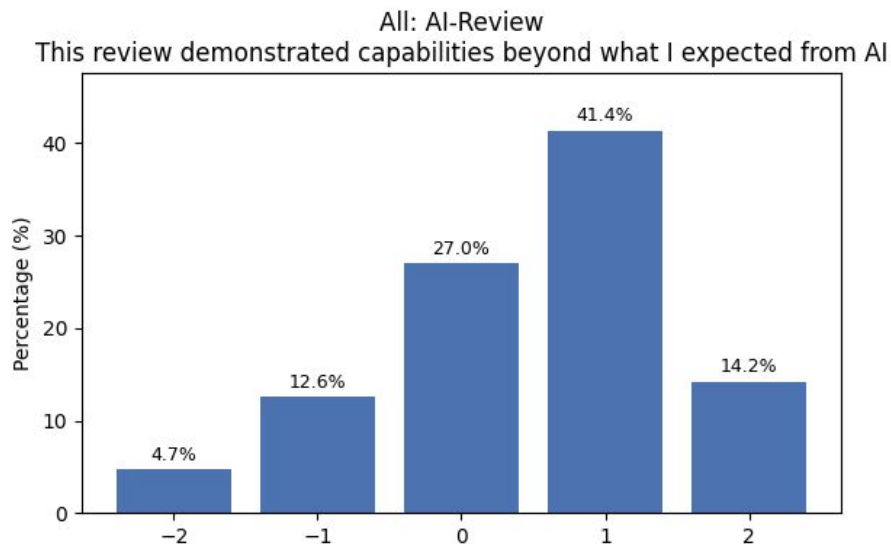
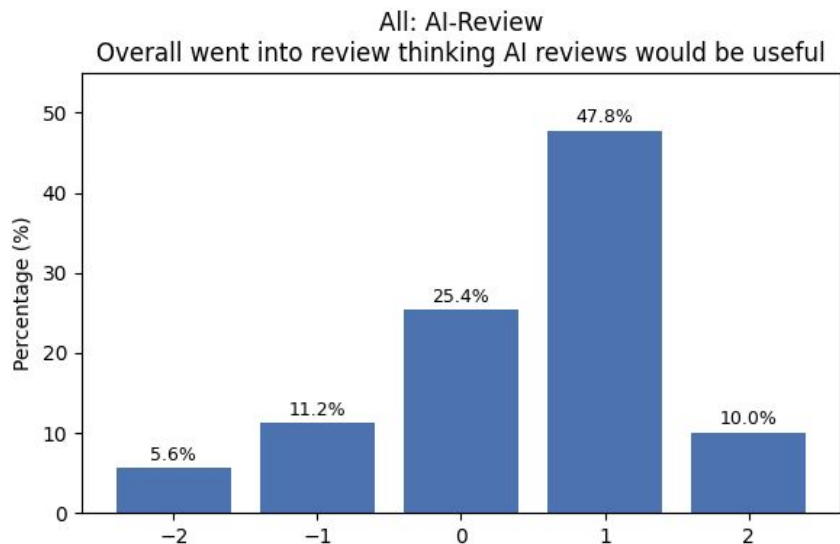
Rejected papers, PC+SPC+AC: AI-Review
This review was a thorough review for this conference



Human Reviews

AI Reviews

Prior Impressions Towards AI In Reviewing



Core Questions: Review and AI

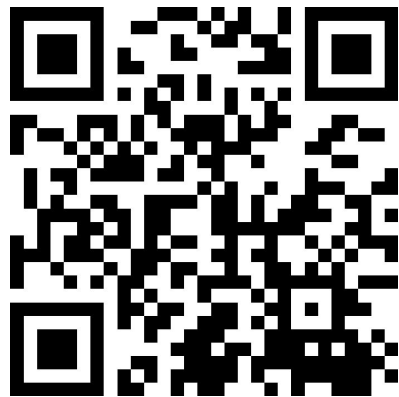


- How can we do a good job for reviewing with so many papers?

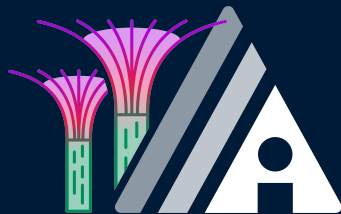
Now and in the future?

- How can we build processes that will continue to do a good job as growth continues?
- What can and should be the role of AI tools?

Review feedback, Scholarly integrity, Paper-reviewer matching ...



Post questions
here (Slido)



The AAAI-2026 AI-Assisted Peer Review Pilot Program

Joydeep Biswas, Sheila Schoepp, Gautham Vasan, Anthony Opipari, Arthur Zhang, Zichao Hu
Chad Jenkins, Matt Taylor, Peter Stone, Kiri Wagstaff, Matt Lease

Panel and Q&A

Joydeep Biswas
The University of Texas at Austin
Associate Program Chair

Odest Chadwicke Jenkins
University of Michigan
Program Co-Chair

Matthew E. Taylor
University of Alberta
Program Co-Chair

Kevin Leyton-Brown
University of British Columbia
Conference Committee Chair