

Rebooting AI Research Publications Strategy: Continuing the Vardi-Halpern discussion, and how to address reviewer burden

Vaishak Belle
University of Edinburgh, UK
vaishak@ed.ac.uk

Abstract

The societal and technological impact of AI is likely to be profound, garnering unparalleled attention on social media and news outlets. There is an urgent need to understand precisely how the recent advancements in AI will impact various fields, ranging from education to the biological sciences. Unfortunately, the dissemination of AI research is currently facing massive bottlenecks.

AI papers are getting submitted to conferences in large numbers, and program committees are struggling. Conference venues in computer science are traditionally treated as archival publications. But as argued by Moshe Vardi, although conferences promise predictability, in terms of submission and decision timing, this is an illusion as papers get bounced from one venue to another until they arbitrarily find a reviewer pool – aligned in spirit and mind perhaps – which accept the paper. Joseph Halpern has argued for some changes within ACM, based on a two-step process. However, both viewpoints ignore the challenge of finding appropriate reviewers for journal submissions, and get timely reports. Besides, conference venues provide an excellent model for dissemination of quickly developing areas.

We advocate a comprehensive strategy consisting of three parts, all supported by a two-step process: (a) conferences for urgent reports, (b) public-facing reviews (or at least accessible to handling editors) for rejected papers, and (c) a timely review process for journals, inspired by but hopefully not as formidable as, the tenure approval process, where authors seek out confirmed reviewers.

1. A conference culture

The societal and technological impact of AI is likely to be profound, garnering unparalleled attention on social media and news outlets [8, 7, 1]. There is an urgent need to understand precisely how the recent advancements in AI will impact various fields, ranging from education to the biological sciences. Unfortunately, the dissemination of AI research is currently facing massive bottlenecks.

AI papers are getting submitted to conferences in large numbers [2, 4, 5, 9, 12, 13], and program committees are struggling.¹ Both large general and specialist conferences in AI, such as AAAI, IJCAI and NeurIPS, get close to 7000 submissions, with acceptance rates of 20% or lesser.² As argued by Moshe Vardi:³

How did we get here? Back in the 1960s, journals were slow, while the field was new and needed to move fast. Conferences offered a solution: Present a preliminary version in a conference with a six-month submit-decide-present cycle, get feedback and credit, and then publish an archival version in a journal. Program committees then did not review papers; they selected papers for the program. Over time, the selection process became a review process: decision notices became decision notices with some feedback, feedback evolved into reviews, and reviews eventually led to rebuttals. Furthermore, conferences have evolved from being a venue for preliminary publication to, in practice, a venue for archival publication. Also, conference-program time pressure, led to selectivity, which led to prestige. As a result, computing-research conferences today are clearly the preferred venue for archival publication.

¹ Although the thrust of this paper might apply more generally to computer science conferences, I do not know of other areas that are getting submissions in as high numbers as the AI community. Here by AI community, I include sub-areas such as pattern recognition and natural language processing.

² <https://neuripsconf.medium.com/what-we-learned-from-neurips-2019-data-111ab996462c>

³ <https://cacm.acm.org/magazines/2021/1/249441-reboot-the-computing-research-publication-systems/fulltext>

But as widely recognized, page limitations (sometimes artificial given that proceedings are almost completely virtual these days) along with the quick review process means that conference publications may not get reviewed for the same type of rigor as journals. They are celebrated for their novelty and the urgency of the results, but detailed discussions on assumptions and mathematical details are regularly omitted, either for space reasons, or because the thrust of the paper does not lie in those details. The larger community, for example, may not care about such mathematical intricacies, so authors omit them altogether. Reviewers do not bother (but neither are they obligated) to check appendices, where some of these intricacies may be spelt out, but even here, proof sketches and the like are standard fare.

The rigor aspect is one dimension of the problem. The arbitrariness of the review process another matter altogether. Numerous community experiments have suggested,⁴ among other things,⁵ that the review process is broken.⁶

Why are conferences still popular then? It is an opportunity for scientists to gather, but this is also true in many other fields too, such as mathematics,⁷ where reasonable abstracts are accepted for conferences, but the hard work and rigorous peer review happens in journals. The second reason is the quick review turnaround for conferences, but as noted by Vardi:

Probably the strongest arguments in favor of conferences are their predictability, in terms of submission and decision timing, and their function as venues for community building. But the promised predictability is an illusion, as papers bounce from conference to conferences until they finally find a home.

We suspect when conference papers are rejected from a major AI conference such as NeurIPS, authors target less prominent AI conferences. We have not been able to find any concrete data on such behaviors, but the open invitation on most call-for-papers that have submission dates slightly after notifications are sent out for a previous major AI conference certainly confirms this. AAAI-2023 allows for a fast track process for NeurIPS rejects, for example.⁸ But the pool of reviewers is usually the same. This leads to an artificial ranking of conferences, that is, getting a paper accepted at one venue versus another becomes a matter of prestige. But clearly this is not helpful, and burdens reviewers unnecessarily. A scientifically “weak” paper can bounce around venues, until it falls into the hands of reviewers who are generous enough to consider it significant and accept it.⁹ Many older conferences, such as IJCAI, AAAI and KR, might be usurped by newer conferences such as AAMAS, ICAART, and ICLR, but this should be based on the size of the conference and community building, and not on arbitrary notions of stronger versus weaker venues, especially when the editorial board, reviewers, and keynote speakers are often shared or essentially the same!

2. Journals the answer?

So what is the solution? Are journals the answer? As Joseph Halpern argues:¹⁰

the median time from submission to final acceptance for papers published in 1997 was 20 months.

If the authors are able to identify the handling editors directly, and with a two-step review process, it could be shortened. Halpern argues:

a two-step review process, with a quick first review to see if a paper meets JACM’s high standard and a more in-depth review if it does,¹¹ has allowed authors whose papers do not quite meet the standards to submit to another journal without incurring the penalty of spending a lot of time in review with us.

⁴<http://inverseprobability.com/2014/12/16/the-nips-experiment/>

⁵<https://cacm.acm.org/blogs/blog-cacm/181996-the-nips-experiment/fulltext>

⁶<https://cacm.acm.org/blogs/blog-cacm/248824-how-objective-is-peer-review/fulltext>

⁷https://www.ams.org/about-us/governance/committees/Statement_ratesofpublication.pdf

⁸<https://aaai.org/Conferences/AAAI-23/neurips-fast-track-submissions/>

⁹Note that we are not making judgements about what constitutes as weak or strong papers, but trust the reviewers to assert this. Of course, there is significant debate on the role of anonymity [11], but we will not comment on that. We are assuming that the review process is well-established, and that rigorous and novel work gets accepted, and so the issue we concern ourselves is really about identifying how to address scalability and avoid overhead in reviewing.

For example, it is entirely reasonable for a venue on, say, “logical foundations of AI” to reject a paper on, say, Rao particle simulations because of fit, or because it does not quite prove as many theorems on logical aspects as expected for a paper in this venue. But this is a reject based on theme and mathematical thrust, or perhaps because the paper’s results are incorrect or incrementally improves on some other author’s paper.

¹⁰<https://dl.acm.org/doi/pdf/10.1145/278298.278301>

¹¹<http://www.acm.org/jacm/Refereeing.html>

Optimistically, it might be possible to bring it down to a year, Halpern suggests. But a year is a long time to wait to ultimately learn that the paper is rejected. Journal reviews, in their current form, still take too long even if a paper is not desk rejected: reviewers often do not respond quickly, and do not submit reviews in time, and this makes the process frustrating [10].

Of course, things have certainly improved since Halpern's observations. These days, Elsevier journal websites, for example, proudly display the number of weeks before an initial review, and these numbers are usually small for most journals. But the reality is that this period is very unpredictable, simply because most researchers have already committed to a full reviewer load at multiple top conferences, including AAAI, IJCAI, ICML, NeurIPS, ECAI, and ICAPS, and the same pool of researchers are often asked for multiple such conferences in a year.

3. A diverse strategy

We think the best way to get out of this quagmire is to upgrade both conference and journal options for authors. This is inspired by many recent endeavors, including eLife ending accept/reject decisions following peer review,¹² an analogous version of which might be worth considering for CS, and certainly AI. How would this work? Here are a few suggestions:

- **Conferences continue to exist.** Completely substituting AI conferences is not a viable suggestion, our cultures are built from ad hoc gatherings at conferences, where a select set of publications are presented and discussed. There is no need to completely abandon them.

But perhaps the burden placed on reviewers can be reduced. For example,

- **Desk reject.** Desk rejection for every paper not passing a sensible sanity check: not plagiarized, not incremental over some other author's work, and so on.
- **Benchmarks.** Results reporting on breakthroughs on benchmarks and competitions are clearly of interest wrt the technology of AI, so provided they pass the sanity check, they could be accepted.
- **Novel theory.** Results on novel theoretical models could also be accepted, provided they pass the sanity check. But given the duration for reviewing, it is unreasonable to expect reviewers to carefully check every mathematical intricacy. They can be conditionally accepted, but reviewers will be given considerably more time, say until the camera ready deadline, to process the details and approve them.
- **Applications.** Papers on exciting applications could be accepted but are not obliged to regurgitate theoretical details from other papers just to "boost" the rigor of the paper. They can be short or long as appropriate without superfluous details.

Most importantly, conferences can continue to stipulate page limits because it reduces the burden on the reviewer, and provided all the necessary details can be discussed within the page limit, the author may prefer conference venues.

- **Rejected papers.** Rejected papers are of many types.
 - **Non-sensical material.** Papers with obviously wrong results, or simply non-sensical material,¹³ should have their reviews accessible to the program committee of every major AI conference. Software such as the Toronto paper matching system¹⁴ could be used to ensure authors do not simply change the title of the paper and a few superficial details and resubmit. If this process is automated, the program committee would never even see or bid on such papers.
 - **Mistaken results.** If the authors have simply made a mistake, which was not deliberate (to err is to be human), then the authors could be given an opportunity to fix and resubmit to the same conference for the next year. They could be strongly discouraged from choosing some other venue just because it has an earlier submission date. After all, they chose the present venue for a reason.
 - **Rigorous but not relevant.** A paper may be strong, but simply not appropriate for the venue. The author, in this case, might not object to the review (and the author's rebuttal) being made public: after all, the program committee thought the paper was technically strong, but simply could not work out its impact on their community. A paper on Cantor's cardinality of the reals [3] does not quite belong in a conference on social robotics, unless perhaps the author has done something clever like show the probabilistic models used by the roboticist are not computable [6].

¹²www.prnewswire.co.uk/news-releases/elife-ends-acceptreject-decisions-following-peer-review-301658712.html

¹³For example, see post and comments in: <http://blog.computationalcomplexity.org/2009/01/so-you-think-you-settled-p-verus-np.html>

¹⁴<https://mila.quebec/en/publication/the-toronto-paper-matching-system-an-automated-paper-reviewer-assignment-system/>

- **Low significance.** We are not in a position to define what low significance means, as this is subjective. Provided it has passed the sanity check (not plagiarized, and not incremental), perhaps low significance can mean that the author has produced a new body of work, but the results have not been explored to an acceptable depth, and so the work is not mature. This is an entirely reasonable stance, and in this case too, the review could be made public, and authors can follow up with a new version with all the fixes if they should so choose.¹⁵

Then there is the issue of author anonymity [11]. We do not really see anyway to solve reviewing burdens if we do not de-anonymize author names. But we would go a step further and say the reviewers should also be made public. Frontiers, for example, makes the reviewers public after acceptance.¹⁶ If there was some need to anonymize reviewers, say to avoid public or professional backlash, perhaps a random selection from the program committee could be added to the set of actual reviewers. The author only knows that, say, three of the mentioned six actually reviewed their work. (Assuming for the sake of discussion that papers get three reviews totally.)

If conferences are going to be such a success, why opt for journals? For the right reasons, we think. For example:¹⁷

- **More pages.** Sometimes carving out the work might just be 20-30 pages long, and the authors can do this without needlessly worrying about the 8 page limit in most AI conferences. Journals may place an upper bound on page limits (e.g., only allowing articles to have a maximum of 60 pages, unless exceptions are granted), of course, simply to allow reviewers to evaluate the material in a reasonable period of time.
- **Niche inquiry.** Although we do not subscribe to this view, reviewers often assess the fit of a rigorous paper for the general community. If conferences make such stipulations, then a rigorous but nice inquiry might be better placed in a journal for the appropriate community. The reviewers might recommend such papers by default, much like the journal track found at many conferences these days.¹⁸
- **Unifying multiple papers.** Multiple conference submissions might deserve a unified treatment to make the subject accessible.
- **Significant expansion of conference submission.** Improvements and newer results over a previously accepted conference paper might deserve a unified treatment.

But if that is the case, how do we address the duration for journal review?

Inspired by but hopefully not as formidable as the tenure approval process, we think one way to shorten the duration needed for the reviews as well as the burden placed on handling editors is the following.¹⁹ We could expect the authors to submit at least six reviewers from the area and another six from slightly outside the area. The idea being that the handling editor will likely choose two from the area and perhaps one from the non-area list. The authors would be expected to justify how the reviewers they have chosen are not conflicts of interest, and this will be made public if the paper is accepted. In fact, after the papers have been peer reviewed and made public, the reviews and the set of reviewers will also be made public. (Thus the authors are in the dark in terms of who exactly reviewed their work, only knowing some among their list were chosen.)

Moreover, authors would be expected to first chase up the reviewers and get confirmation that they are actually willing to do the reviews in the selected period of time. In this case when there is a delay in getting a review back, its the authors failing, and the handling editor is not necessarily to be blamed. After consecutive prompts from the handling editor, if the reviewer still does not get back to them, the handling editor can look to the next person on the appropriate list. And so on, until the list is completed. If none of the identified reviewers respond, the work is rejected but purely on grounds of reviewers not responding.

This strategy is not without its problems. In case the authors are doing substantial work in an interesting area where the reviewers are motivated to do their job on time, and if these papers are rigorous and thorough, then they would get accepted quickly. But conversely, papers in slightly esoteric areas where the reviewer pool is small or not active, papers might be in a limbo. This is unfortunate, but peers are essential for peer review, so there might not really be a way to get around this.

¹⁵Reviews being made public is very much in the spirit of OpenReview (<https://openreview.net>), but our concerns and suggestions are much broader.

¹⁶<https://www.frontiersin.org/about/peer-review>

¹⁷See, for example, criteria for JAIR: <https://www.jair.org/index.php/jair/about/submissions>

¹⁸<https://ijclr22.doc.ic.ac.uk/papers/index.html>

¹⁹The evaluation strategy for project proposals, which is typically based on some reviewers identified by the applicants and others chose by the funding panel, is very similar in spirit, and is also a source of inspiration.

4. Conclusions

AI papers are getting submitted to conferences in large numbers, and program committees are struggling. Page limits and short durations does not allow in all instances for quality reviews. So we need to balance speedy publishing of new developments, where conferences might be appropriate, with rigorous reviewing for more elaborate material. Both might then be seen as equally prestigious: conferences for applications, benchmarks and noteworthy technical insights, and journals for deep, unifying work. We avoid reviewing burdens by making reviews public, and discouraging resubmissions from one venue to another (unless justified because of distinction in theme and emphasis from present venue). Finally, handling editors for journals approach a list identified to be no-conflict by the authors which will be made public on acceptance. A two-step review strategy needs to be in place everywhere, to avoid a detailed look at work that is wrong, incomplete or not fitting.

We do not wish to conclude by stating that we have identified and addressed the problem. Instead, we propose a strategy based on academic processes that involve external reviewers, which nonetheless get reviewed on timely. Therefore, we hope that this marks the beginning of a discussion that could ultimately result in the timely publication of technically rigorous and interesting results from the scientific community. Given the rapid pace at which artificial intelligence is evolving, timely publication is crucial.

References

- [1] Arrieta, A. B., Rodríguez, N. D., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., and Herrera, F. (2019). Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *CoRR*, abs/1910.10045.
- [2] Boudry, C., Al Hajj, H., Arnould, L., and Mouriaux, F. (2022). Analysis of international publication trends in artificial intelligence in ophthalmology. *Graefe's Archive for Clinical and Experimental Ophthalmology*, 260(5):1779–1788.
- [3] Cantor, G. (1874). On a property of the class of all real algebraic numbers. *Crelle's Journal for Mathematics*, 77(1874):258–262.
- [4] Chan, J., Liao, C., Mann, A., Kapp, D., and Mysona, D. (2020). 103 artificial intelligence publication trends in reproductive cancers—who is being left behind?
- [5] Cockburn, I. M., Henderson, R., and Stern, S. (2018). The impact of artificial intelligence on innovation: An exploratory analysis. In *The economics of artificial intelligence: An agenda*, pages 115–146. University of Chicago Press.
- [6] Huang, D. and Morrisett, G. (2016). An application of computable distributions to the semantics of probabilistic programming languages. In *European Symposium on Programming*, pages 337–363. Springer.
- [7] Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. (2017). Human Decisions and Machine Predictions*. *The Quarterly Journal of Economics*, 133(1):237–293.
- [8] Liao, Q. V., Gruen, D., and Miller, S. (2020). *Questioning the AI: Informing Design Practices for Explainable AI User Experiences*, pages 1–15. Association for Computing Machinery, New York, NY, USA.
- [9] Matheny, M., Israni, S. T., Ahmed, M., and Whicher, D. (2019). Artificial intelligence in health care: The hope, the hype, the promise, the peril. *Washington, DC: National Academy of Medicine*.
- [10] Nguyen, V. M., Haddaway, N. R., Gutowsky, L. F., Wilson, A. D., Gallagher, A. J., Donaldson, M. R., Hammerschlag, N., and Cooke, S. J. (2015). How long is too long in contemporary peer review? perspectives from authors publishing in conservation biology journals. *PloS one*, 10(8):e0132557.
- [11] Pontille, D. and Torny, D. (2014). The blind shall see! the question of anonymity in journal peer review. *The Question of Anonymity in Journal Peer Review*.
- [12] Shiferaw, K. B., Waltemath, D., and Zeleke, A. (2022). Disparities in regional publication trends on the topic of artificial intelligence in biomedical science over the last five years: A bibliometric analysis. *Studies in Health Technology and Informatics*, 294:609–613.
- [13] Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., et al. (2022). Artificial intelligence and life in 2030: the one hundred year study on artificial intelligence. *arXiv preprint arXiv:2211.06318*.