Towards A Sustainable Future for Peer Review in Software Engineering

Esteban Parra
esteban.parrarodriguez@belmont.edu
Belmont University
Nashville, Tennessee, USA

Sonia Haiduc
shaiduc@fsu.edu
Florida State University
Tallahassee, Florida, USA

Preetha Chatterjee
preetha.chatterjee@drexel.edu
Drexel University
Philadelphia, Philadelphia, USA

Ramtin Ehsani
ramtin.ehsani@drexel.edu
Drexel University
Philadelphia, Philadelphia, USA

Polina Iaremchuk
polina.iaremchuk@bruins.belmont.edu
Belmont University
Nashville, Tennessee, USA

Abstract

Peer review is the main mechanism by which the software engineering community assesses the quality of scientific results. However, the rapid growth of paper submissions in software engineering venues has outpaced the availability of qualified reviewers, creating a growing imbalance that risks constraining and negatively impacting the long-term growth of the Software Engineering (SE) research community.

Our vision of the Future of the SE research landscape involves a more scalable, inclusive, and resilient peer review process that incorporates additional mechanisms for: 1) attracting and training newcomers to serve as high-quality reviewers, 2) incentivizing more community members to serve as peer reviewers, and 3) cautiously integrating AI tools to support a high-quality review process.

CCS Concepts

• **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

Keywords

FutureSE, peer-review, juniorPC, responsible AI use, AI ethics

ACM Reference Format:
Esteban Parra, Sonia Haiduc, Preetha Chatterjee, Ramtin Ehsani, and Polina Iaremchuk. 2018. Towards A Sustainable Future for Peer Review in Software Engineering. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (ICSE-Future'26).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Peer review is the backbone of trust, legitimacy, and collective knowledge-building in software engineering. Similarly to other disciplines, the decision on whether to accept a paper for a journal or a conference lies with editors and Program Committee (PC) chairs, respectively [12]. However, due to the volume of submissions and to avoid biases in the decision-making process, editors and chairs make their decision primarily based on the reviews provided by a set of expert reviewers [12].

Reviewers are trusted individuals invited to provide professional reviews for papers based on their expertise and standing in the research community. Generally speaking, reviewers are assigned a set of papers and have the following core responsibilities for each paper: reading the paper, assessing the paper's quality and fit for publication in the venue using various criteria (e.g., novelty, soundness, relevance, presentation, etc.), giving a recommendation for acceptance or non-acceptance of the work, and providing clear and constructive feedback to the authors.

Despite reviewers' work being essential for the software engineering research community, reviewing research papers is unpaid, volunteer work performed by members of the community in addition to their multiple tasks as researchers and/or educators [18]. Submission volumes have grown faster than reviewer availability, leading to high reviewer load, with some members of the community conducting over 20 reviews per year for conferences alone [8]. Furthermore, reviewing is a time-consuming task; for example, 88% of reviewers spend over 2 hours to read and review a single journal paper [8]. This can lead reviewers to work on the weekends or late in the evenings [8]. At the same time, reviewing for the main tracks of larger conferences can be a year-long commitment, limiting reviewer availability for other venues or tracks.

Figures 1 and 2 show the changes in the last five years with respect to the size of the program committee and number of submissions to the main tracks of some of the largest software engineering conferences ranked A* and A (i.e., ICSE, FSE, ASE, ICSME, MSR, and SANER).

Figure 2 is the number of main track submissions that underwent peer review (i.e., without desk rejections). The figure shows that in the last two years, the number of submissions has increased significantly. To account for this increase in submissions, the largest conferences have also increased their PC size. However, even with the increased PC size, the reviewer workload has often increased given that each paper is usually assigned at least three reviewers. For example, for ASE, the size of the main track PC has more than doubled between 2020 and 2025 (from 148 to 326 PC members),

ICSE-Future'26, Woodstock, BR

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

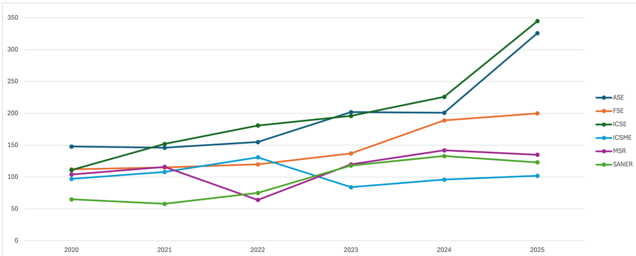

Figure 1: PC Size Over Time

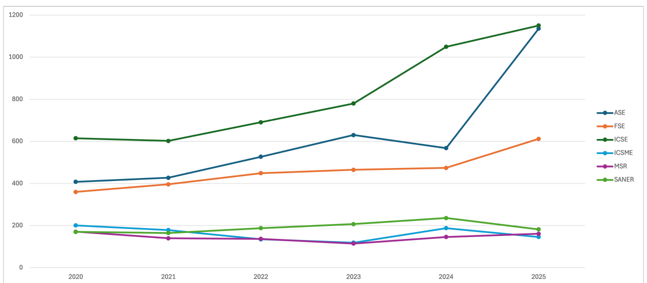

Figure 2: Submissions Over Time

while the number of main track submissions has nearly tripled (408 in 2020 vs 1136 in 2025), leading to a workload increase from 8.27 to 10.45 papers per PC member on average. Furthermore, many of the community members serve on the PC of multiple conferences or on the PCs of multiple tracks of the same conference, leading to an overall increasing review load.

As it stands, the discrepancy between the growth in submissions and the availability of reviewers is not sustainable and risks harming the quality of the reviewing process and the potential for growth of the SE community. We envision a future of SE research in which quality peer review is made sustainable by expanding the pool of high-quality reviewers through scalable training, retaining, and incentivizing both established community members and newcomers to serve as peer reviewers.

2 What Is Working

The feedback provided by reviewers on the quality of papers is incredibly valuable, as it allows authors to expand and strengthen their work before it is published. Even in cases of non-acceptance, the feedback provides actionable steps from experienced researchers on potential critical issues that, when addressed, should improve the paper for future submissions. Furthermore, the variety of tracks with different review criteria, such as the Reproducibility Studies and Negative Results at SANER[1] and ICSME[2], increases venue accessibility to a wider range of work and is a valuable tool for growing the community.

[1]https://conf.researchr.org/track/saner-2026/saner-2026-reproducibility-studies-and-negative-results-rene-track
[2]https://conf.researchr.org/track/icsme-2026/icsme-2026-replication-and-negative-results

2.1 Distinguished Reviewer Awards

Peer review is viewed by members of the community as a shared form of quality control. Although the reasons for accepting to be a reviewer vary, many reviewers accept this role from a sense of professional duty.

The main mechanism used by most software engineering research conferences to recognize the reviewers' labor is through the Distinguished Reviewer Award. This award is presented to a subset of reviewers who have made the greatest contributions to the review process [18]. This type of award represents a valuable recognition of a researcher's efforts in the scientific community and an addition to the CV that could support their career goals [18]. Therefore, it can serve as an incentive to participate in the reviewing process. However, given that very few reviewers get this recognition, it should ideally be complemented by other types of incentives for recruiting reviewers.

2.2 Junior/Shadow PC

The 18th edition of the IEEE/ACM International Conference on Mining Software Repositories (MSR'21) established the Shadow PC mentorship program, later renamed Junior PC, as a mechanism to provide opportunities for early-career researchers, namely PhD students, post-docs, new faculty members, and industry practitioners, to learn about and get involved in the academic peer review process, aiming to increase the pool of qualified reviewers [17]. The members of the Shadow/Junior PC are integrated in the peer review process with the members of the main technical track PC using a 2-1 model, where each paper is reviewed by 2 members of the regular PC and 1 member of the Junior PC. This way, Junior PC members have the opportunity to observe and learn from the reviews of more seasoned members of the community and can also receive feedback on their own reviews.

The Junior PC mentorship program continues to be a part of MSR[3] to this day, as it has proven to be an effective mechanism to increase the reviewer pool in the software engineering research community. Furthermore, a similar mentorship program, the Shadow PC, has been incorporated in ICSE[4], providing an opportunity for early-career researchers and graduate students who have not yet served on a technical research track program committee at international SE conferences to learn and contribute to the peer review process in one of the most prestigious venues in the field [19].

3 What Is Not Working

Reviewing can be invisible labor with little career recognition in a "Publish or Perish" research culture, which makes it challenging for the reviewer pool to expand organically.

3.1 Scale and Overload

While peer review is crucial for the community, in practice it is dependent on the community members' workload and availability. To make the process scalable and sustainable, the commonly referenced goal is "to review as much as you are reviewed" [8]. This means that for each paper an author submits, that author would be

[3]https://2026.msrconf.org/track/msr-2026-junior-pc
[4]https://conf.researchr.org/track/icse-2026/icse-2026-shadow-research-track-program-committee

expected to perform three reviews in return (assuming an average of three reviewers per paper). However, due to the disjoint nature of the venues, there is currently no mechanism to track or enforce this practice.

Notably, one of the growing pains in the SE research community is that submission volumes have grown faster than reviewer capacity, as the availability of quality researchers to serve as reviewers for papers in both conferences and journals is limited, leading to frustration from reviewers and authors alike [3, 15]. Figure 3 presents a recent LinkedIn post from a member of the software engineering research community that echoes this sentiment and further highlights potential concerns with respect to the quality of the reviews.

Software engineering conferences like ICSE, FSE, and ASE continue to serve as vital hubs for empirical and technical advances, yet they face mounting tension between quantity and quality.

Submissions have exploded, but reviewer workloads haven't, often forcing hurried, heuristic evaluations that compromise depth, reproducibility, and fairness .
Rigid formatting and track structures discourage bold, interdisciplinary research—favoring incremental, "safe" papers that fit neat templates.

Despite these flaws, conferences remain invaluable for networking and community cohesion—but their long-term scholarly impact depends on embracing transparent reviews, incentive realignment, and open, adaptive processes.

#SoftwareEngineering #PeerReview #ICSE #FSE #ASE #AcademicPublishing #Reproducibility #BiasInScience

Figure 3: A Linked-In Post on Peer Review in SE Venues

Advances in the natural language capabilities of Large Language Models (LLMs) and Generative Artificial Intelligence (Gen-AI) have enabled these tools to also generate research paper reviews [4, 12]. There are growing concerns regarding the practice of some reviewers to generate and submit AI reviews with little or no informed editing or corrections, particularly given the fact that AI tools lack expertise in the field and may even replicate existing biases [4, 14, 16]. In addition, providing a manuscript under review to an AI tool can lead to violating the confidentiality of the peer review process, given that AI tools can "learn" from the content of the manuscript. As a consequence of these concerns, ACM and IEEE currently have policies in place regarding the use of generative AI or LLMs in peer review. These guidelines specify that "uploading any part of a submission to an LLM or other third-party system that does not promise to maintain the confidentiality of that information is not permitted" [2, 9].

Despite existing guidelines limiting the use of AI in the peer review process, recent cases in computer science [14] and other disciplines [13] have brought to the forefront issues regarding the rising proliferation of low-quality reviews produced by AI in peer reviewing. Attributes of low-quality reviews resulting from reviewers' use of AI include: fake/hallucinated citations, suspiciously long and vague feedback, and methodologically incorrect suggestions that do not make sense [14].

Notably close to the SE community is the recent case involving ICLR'26, one of the top machine learning conferences (see Figure 4). An analysis of the paper reviews submitted by PC members revealed that over 15,000 peer reviews were fully AI-generated

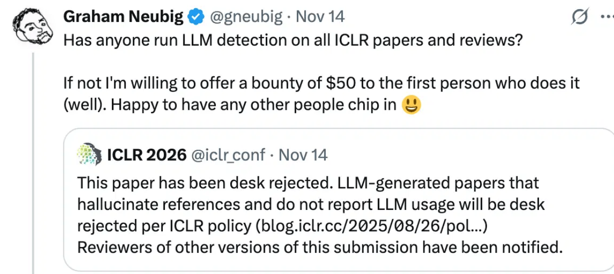

Figure 4: Tweet on AI-generated Papers and Reviews at ICLR

[7, 14]. The ICLR'26 case serves as a *canary in the coal mine* for our community on the risks and potential negative consequences of the rising imbalance between paper submissions and reviewing capacity.

3.2 Barriers to Entry

Reviewing papers for conferences and journals constitutes a form of invisible labor with little career recognition in our "Publish or Perish" research environment. The unpaid, volunteer nature of peer review creates a fundamental asymmetry: reviewers donate substantial time and expertise to the scientific community, yet this contribution remains largely unrecognized and unrewarded within institutional career advancement systems. A recent survey of peer reviewing in software engineering by Ernst et al. [8] showed that early-career researchers prioritize reviewing for venues to determine what makes a paper strong enough to be published at top-tier venues. However, this learning opportunity comes at a personal cost, as the time investment in reviewing directly competes with their own research and publication activities.

At many institutions, reviewing is not meaningfully valued as service within tenure and promotion processes. As a result, researchers are not incentivized to serve as reviewers, as it often does not contribute directly to the advancement of their career goals [8]. This lack of institutional recognition is particularly problematic for early-career researchers (ECRs) who face pressure to maximize their publication output and secure independent funding. In such a competitive landscape, investing significant time in unrewarded service work becomes a luxury that many ECRs cannot afford.

The existing mechanisms for recognizing reviewer contributions are insufficient to address these barriers. While valuable, Distinguished Reviewer Awards are usually awarded to more experienced members of the community, with less than 3% of award recipients having fewer than 5 years of experience [8]. Therefore, these awards function as a form of validation for more established researchers rather than as an incentive for attracting newcomers to the peer-reviewing process.

Finally, existing mentorship initiatives such as the Junior/Shadow PC programs currently have important limitations that restrict their impact. The Shadow PC program, pioneered by MSR'21 and subsequently adopted by ICSE, provides valuable hands-on experience for ECRs. However, these programs are typically limited to a single review cycle and have not been widely implemented across conferences and journals, thus significantly limiting their reach and scalability. As a result, only a small subset of the growing population of ECRs can benefit from these mentorship opportunities, leaving the vast majority without structured guidance on how to conduct high-quality reviews or establish themselves as valued members of the peer review community.

4 A More Sustainable Future of Peer Review in SE

In this section, we provide a set of mechanisms whose adoption can build a more sustainable peer review ecosystem in the SE research community.

4.1 Scalable Training of Junior Reviewers

Several conferences have previously hosted presentations by experienced reviewers from the community about writing high-quality peer reviews. However, these efforts have never been broadly available to the wider community.

A more sustainable future for SE peer review has to include ways of making training for junior peer reviewers more scalable. This could be achieved by creating an online training module on writing high-quality reviews for research papers, akin to the CITI training for Responsible Conduct of Research and Human Subjects Research [5]. The training would include videos from award-winning reviewers from the SE community and PC chairs and would systematically address the various quality aspects of a paper: motivation, novelty, methodology, presentation, etc.; reasons to reject/accept a paper; common do's and don'ts; what to do when you observe low-quality reviews; identifying AI-generated papers; responsible and allowable AI use in peer reviewing; and examples of high-quality and low-quality reviews.

The training could also be complemented by an online platform/forum, where trainees could review some sample published papers or draft papers of their peers and receive feedback on their reviews from fellow trainees, as well as more senior members of the community.

The online course would include a quiz at the end and issue a certificate of completion, and then post the names of the graduating trainees on a public website, along with their affiliation, research expertise, Google Scholar link, reviewing experience, and reviewing availability. In addition, reviewers could update their social media or ORCID profiles to add their certificate of completion as well as their availability, etc. This information could then be used by PC chairs and editors looking for reviewers.

4.2 Responsible Use of AI in the Peer Review Process

As mentioned earlier, the use of Gen-AI by reviewers to produce fully AI-generated reviews of research papers is a growing and worrying issue. Although recent work on the use of agentic peer review tools has shown promising results [6, 10, 11], AI should not be used to replace the reviewers' expertise in the field or to bias their professional opinion about a paper.

Nonetheless, we see value and potential benefits in cautiously and responsibly using Gen-AI as a tool to assist in the peer review-ing process.

First, journal editors and PC chairs should clearly define the allowed and forbidden uses of AI in paper submissions and peer reviews. As an example, the ICLR 2026 chairs allowed authors and reviewers to use AI tools to polish text, generate experiment codes, or analyze results, but required disclosure of such uses and also prohibited AI use that would breach the confidentiality of manuscripts or produce falsified content [14]. Next, editors and PC chairs could use AI tools such as Pangram Labs [14] and EditLens [16] to assess the degree of AI-generated content in both paper submissions and peer reviews. Violations of the established policies should be penalized in order to discourage future incidents. This would benefit the community by reducing the number of low-quality submissions to be reviewed as well as deterring reviewers from using AI to produce their reviews.

Further, after all reviewers turn in their own, human-written reviews, AI could be used to generate an additional, complementary review to be used by reviewers to uncover potential issues that may have been overlooked. Most importantly, the AI-generated review should not be available before all the reviewers have provided their expert reviews and have had an initial discussion about the paper; this is crucial in order to avoid the introduction of bias or undue influence on the reviewers' judgments. AI could also be used to craft starting drafts for meta-reviews, which would then be validated, modified, and expanded by the reviewers to appropriately summarize their discussion and assessment of the manuscript.

Last but not least, in order to adhere to ACM and IEEE guidelines for Gen-AI use during the review process, we would need to have community discussions with these organizations to determine the extent to which Gen-AI use is permitted and enforce the use of enterprise versions of LLMs or other third-party systems that guarantee the confidentiality of that information in order to preserve the confidential nature of the papers being reviewed [2].

4.2.1 Incentives. It stands to reason that more experienced researchers, who received and wrote many reviews, can provide better reviews and thus receive the Distinguished Reviewer Awards more often [8]. However, this dynamic diminishes its effectiveness in attracting new reviewers.

We believe that the following strategies to build upon and complement the existing incentives for reviewers in the community would be beneficial in the llong term:

- Submission requirement: Adding a requirement for authors submitting a paper to the conference/venue to serve as reviewers as well. This would mirror the practice in NLP conferences such as ACL [1]. Authors would have to go through the training module previously mentioned to ensure they have the required know-how to provide high-quality reviews.
- Registration Discount: Providing a small discount for conference registration to members of the community serving as reviewers.
- Reviewer badges: Adding a visible stamp on registration badges that identify members of the program committee and Distinguished Reviewer Award recipients.
- Distinguished Newbie Reviewer Award: Expanding the Distinguished Reviewer Award by reserving a small fraction

of the awards to be exclusively for newcomers in the community.

5 Conclusion

The sustainability of peer review is a critical challenge facing the software engineering research community. In this paper, we present our vision to improve the peer review process our community relies upon and move us towards a more sustainable and rewarding peer review system that supports the continued growth and impact of the community through high-quality reviews.

Realizing this vision of the future of peer review in the SE research community will require coordinated efforts from program committee chairs, journal editors, professional organizations, and researchers in the community, but it is a viable and necessary path to ensure the survival of the peer review process in our community.

References

[1] 2026. Call for Main Conference Papers - ACL 2026: Mandatory Reviewing Workload. https://2026.aclweb.org/calls/main_conference_papers/#mandatory-reviewing-workload. Accessed: 2026-01-09.

[2] Association for Computing Machinery (ACM). 2025. ACM Peer Review Policy FAQ. https://www.acm.org/publications/policies/peer-review-faq. Accessed: 2026-01-06.

[3] Lionel Briand. 2025. *Why do top journals reject good papers?* https://www.linkedin.com/posts/lionel-briand-5082b1_why-do-top-journals-reject-good-papers-activity-7411854245508911106-x1rO LinkedIn post, accessed January 14, 2026.

[4] Alessandro Checco, Lorenzo Bracciale, Pierpaolo Loreti, Stephen Pinfield, and Giuseppe Bianchi. 2021. AI-assisted peer review. *Humanities and social sciences communications* 8, 1 (2021), 1–11.

[5] CITI Program. 2025. Social-Behavioral-Educational (SBE) Foundations. https://about.citiprogram.org/course/social-behavioral-educational-sbe-foundations/. Accessed: 2026-01-06.

[6] Kun Ding, Ying Wang, and Shiming Xiang. 2025. EvoVLMA: Evolutionary Vision-Language Model Adaptation. In *Proceedings of the 33rd ACM International Conference on Multimedia*. 4619–4628.

[7] Bradley Emi. 2025. Pangram Predicts 21% of ICLR Reviews are AI-Generated. https://www.pangram.com/blog/pangram-predicts-21-of-iclr-reviews-are-ai-generated. Accessed: 2026-01-14.

[8] Neil A Ernst, Jeffrey C Carver, Daniel Mendez, and Marco Torchiano. 2021. Understanding peer review of software engineering papers. *Empirical Software Engineering* 26, 5 (2021), 103.

[9] IEEE Access. 2025. Reviewer Best Practices. https://ieeeaccess.ieee.org/reviewers/reviewer-best-practices/. Accessed: 2026-01-06.

[10] Yixing Jiang and Andrew Ng. 2025. paperreview.ai: Stanford Agentic Reviewer. https://paperreview.ai/. Accessed: 2026-01-09.

[11] Weixin Liang, Yuhui Zhang, Hancheng Cao, Binglu Wang, Daisy Yi Ding, Xinyu Yang, Kailas Vodrahalli, Siyu He, Daniel Scott Smith, Yian Yin, et al. 2024. Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *NEJM AI* 1, 8 (2024), AIoa2400196.

[12] Jialiang Lin, Jiaxin Song, Zhangping Zhou, Yidong Chen, and Xiaodong Shi. 2023. Automated scholarly paper review: Concepts, technologies, and challenges. *Information fusion* 98 (2023), 101830.

[13] Miryam Naddaf. 2025. AI tool detects LLM-generated text in research papers and peer reviews. *Nature* 648, — (2025), 256–257. doi:10.1038/d41586-025-02936-6 Online ahead of print; accessed 2026-01-14.

[14] Miryam Naddaf. 2025. Major AI conference flooded with peer reviews written fully by AI. *Nature* 648, 8093 (2025), 256–257. doi:10.1038/d41586-025-03506-6 Published 27 November 2025; accessed 2026-01-14.

[15] Mahan Tafreshipour. 2025. *Challenges in software engineering conferences: quantity vs quality, peer review, and reproducibility.* https://www.linkedin.com/posts/mahan-tp_softwareengineering-peerreview-icse-activity-7349359398864998400-nkMI LinkedIn post, accessed January 14, 2026.

[16] Katherine Thai, Bradley Emi, Elyas Masrour, and Mohit Iyyer. 2025. EditLens: Quantifying the Extent of AI Editing in Text. arXiv:2510.03154 [cs.CL] https://arxiv.org/abs/2510.03154

[17] Patanamon Thongtanunam, Ayushi Rastogi, Foutse Khomh, Serge Demeyer, Meiyappan Nagappan, Kelly Blincoe, and Gregorio Robles. 2021. Shadow program committee initiative: Process and reflection. *ACM SIGSOFT Software Engineering Notes* 46, 4 (2021), 16–18.

[18] Bianca Trinkenreich. 2025. *Reviewing academic papers is challenging yet deeply rewarding.* https://www.linkedin.com/posts/biancatrink_reviewing-academic-papers-is-challenging-activity-7307745057300697088-t8pM LinkedIn post, accessed January 14, 2026.

[19] Delvin Varghese, Hafeni Mthoko, Jessica Watterson, Pranav Kulkarni, Dan Richardson, Shaimaa Lazem, and Patrick Olivier. 2024. Shadow Program Committee: Designing for Diversity and Equity within Academic Communities. *ACM J. Comput. Sustain. Soc.* 2, 1, Article 2 (Jan. 2024), 24 pages. doi:10.1145/3630743