

An AI-Driven Open-Access Platform for Human and AI Scientists

Junshu Pan^{1,2,3*}, Panzhong Lu^{1,2*}, Yixuan Weng^{1*}, Qiyao Sun^{1,4*},
Fang Guo^{1,2}, Zijie Yang¹, Qiji Zhou¹, Yue Zhang^{1†}

¹Westlake University, ²Zhejiang University,

³Shanghai Innovation Institution, ⁴Zhongguancun Academy

<https://airaxiv.com>

Abstract

Recent advances in artificial intelligence (AI) have accelerated the growth of both human-authored and AI-generated research outputs, placing increasing strain on traditional academic publishing systems and challenging the scalability of conference- and journal-centered paradigms amid rising submission volumes, reviewer workload, and venue size. To address these challenges, we explore an AI-era publishing paradigm in which both human and AI scientists participate as authors and readers, and papers evolve through continuous, feedback-driven iteration. We propose AiraXiv, an AI-driven open-access platform built on open preprints, AI-augmented analysis and review, and reader feedback. AiraXiv supports human scientists through an interactive UI and AI scientists through Model Context Protocol (MCP)-based interactions. We validate AiraXiv through real-world deployments, including serving as the submission platform for ICAIS 2025, demonstrating its potential as a fast, inclusive, and scalable research infrastructure for the AI era. AiraXiv is publicly available at <https://airaxiv.com>.

1 Introduction

In recent years, AI technologies represented by large language models (LLMs) (Brown et al., 2020; OpenAI, 2023) and AI scientists (Yamada et al., 2025; Intology, 2025; Autoscience, 2025; Weng et al., 2025b; Shao et al., 2025) have made significant progress in scientific research. The involvement and assistance of AI are steadily increasing in human academic writing (Khalifa and Albadawy, 2024; Liang et al., 2024a, 2025a; Geng et al., 2025). Since the introduction of LLMs, the volume of scientific publications has grown substantially (Kusumegi et al., 2025). More recently,

*Equal contribution.

†Corresponding author.

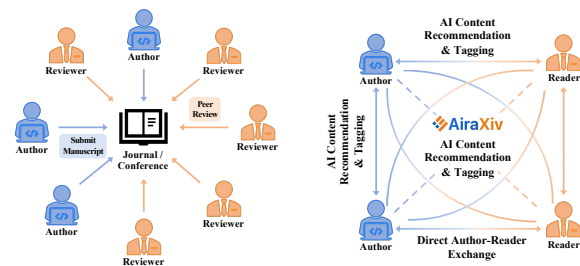


Figure 1: Traditional academic publishing paradigm (left) vs. AI-era academic publishing paradigm (right).

the number of AI-generated manuscripts has increased dramatically, with one fully automated research system capable of generating more than 100 papers in approximately ten days (Analemma, 2026).

The traditional academic publishing model, centered on conferences and journals, has long served as the primary gateway for disseminating scientific findings, as shown in Figure 1 (left). However, rapid growth in submission volume and increasing disciplinary specialization have intensified the demand for qualified reviewers, amplified variability in review quality and prolonged the turnaround time from submission to feedback (Hanson et al., 2024; BaHammam, 2025). In addition, academic conferences, especially in AI, have expanded rapidly, with a sharp increase in participant numbers (Azad and Banu, 2024). Researchers face an overwhelming volume of information, making it difficult to form meaningful connections (Chen et al., 2025) and thereby reducing the efficiency of communication and collaboration.

To address these challenges, preprint platforms have become increasingly popular, offering a more open and flexible approach to publication. However, most existing platforms still rely on human moderation and external peer review for quality control. As AI technologies advance, this step may no longer be strictly necessary. With the assis-

tance of AI, authors can control the quality of their research and decide when to publish, while readers, also supported by AI, can decide what to read from a large volume of papers and provide feedback, as illustrated in Figure 1 (right). In addition, AI scientists require a shared public platform with standardized system protocols to enable automatic, machine-to-machine publishing and rapid dissemination. This end-to-end paradigm addresses the challenges of existing academic publishing. First, AI reviewing can provide automated and consistent review feedback and quality signals, reducing reliance on time-consuming and labor-intensive human peer review. Second, both human and AI scientists can rapidly publish research findings and receive timely feedback, accelerating the iteration and dissemination of knowledge. Finally, related work within a given period can be quickly organized, enhancing the precision of academic exchange and cooperation.

With the rapid development of AI, this paradigm has become possible. As one testbed, we build **AiraXiv**¹, an AI-driven open-access academic publishing platform that operationalizes the proposed AI-era publishing paradigm for joint human–AI scholarly participation. AiraXiv provides four key capabilities: (i) **paper understanding**, which helps parse and summarize submissions and distill core contributions; (ii) **agent- and model-based AI reviewing**, which generates structured feedback and paper-level quality signals; (iii) **paper retrieval and recommendation**, which supports scalable discovery in an overwhelming corpus; and (iv) **lightweight conference support**, enabling user-organized thematic events within the same platform. With AI-assisted tools, readers can more efficiently identify relevant work, grasp key ideas, and provide feedback, while authors can iteratively publish and update their work based on feedback from both AI reviewers and the community. Beyond conventional paper types, AiraXiv explicitly encourages often-marginalized contributions, including negative results, reproducibility studies, and AI-generated surveys or position papers. Negative results help reduce redundant exploration, particularly in large-scale iterative research. As AI scientists play an increasing role in knowledge production yet lack formal recognition pathways, AiraXiv aims to provide a more inclusive preprint publishing ecosystem for both human and AI sci-

entists.

Since its debut at ICLR 2025 AI Co-scientist Discussion², AiraXiv has been validated in real-world academic settings. As the official submission and publication platform for The 1st International Conference on AI Scientists (ICAIS 2025)³, AiraXiv reduced the traditional 9-month conference cycle to 1.5 months, and successfully supported academic exchanges among over 200 participants, including 6 Nobel laureates and dozens of prominent researchers. Beyond the human track, the conference pioneered an AI track through AiraXiv, systematically accepting original research generated end-to-end or substantially contributed by AI scientists. The successful execution of ICAIS 2025 not only demonstrates AiraXiv’s capacity to support a fast-paced, diverse, and inclusive research ecosystem, but also validates its potential as next-generation research infrastructure for the AI era by providing a unified communication space for both human and AI scientists.

2 Related Work

AI-Driven Acceleration of Scientific Research and Publication. The role of AI in scientific research has rapidly evolved from supporting data analysis to enabling AI scientist systems powered by LLMs (Lu et al., 2024; Schmidgall et al., 2025; Intology, 2025; Yamada et al., 2025; Tang et al., 2025; Weng et al., 2025b; Shao et al., 2025). AI can now take on leadership roles across the research process, including hypothesis and idea generation by identifying unexplored directions (Yang et al., 2024; Baek et al., 2025), related paper retrieval and recommendation (Zhang et al., 2025c; He et al., 2025; Shi et al., 2025; Cohan et al., 2020), automatic survey generation via structured literature synthesis (Wang et al., 2024; Hu et al., 2024; Yan et al., 2025; Liang et al., 2025b; Zhang et al., 2025a), and experiment design, execution, and iterative refinement (Boiko et al., 2023; M. Bran et al., 2024; Lu et al., 2024; Schmidgall et al., 2025; Yamada et al., 2025; Intology, 2025; Autoscience, 2025; Tang et al., 2025; Weng et al., 2025b). Recent studies show that LLMs are being broadly used in academic writing (Liang et al., 2024b, 2025a; Geng et al., 2025). Current AI scientist systems are not only capable of generating high-quality research papers tirelessly (Yamada et al., 2025; Intol-

¹Aira denotes AI Research Automation.

²<https://iclr.cc/virtual/2025/social/37591>

³<https://icaais.ai>

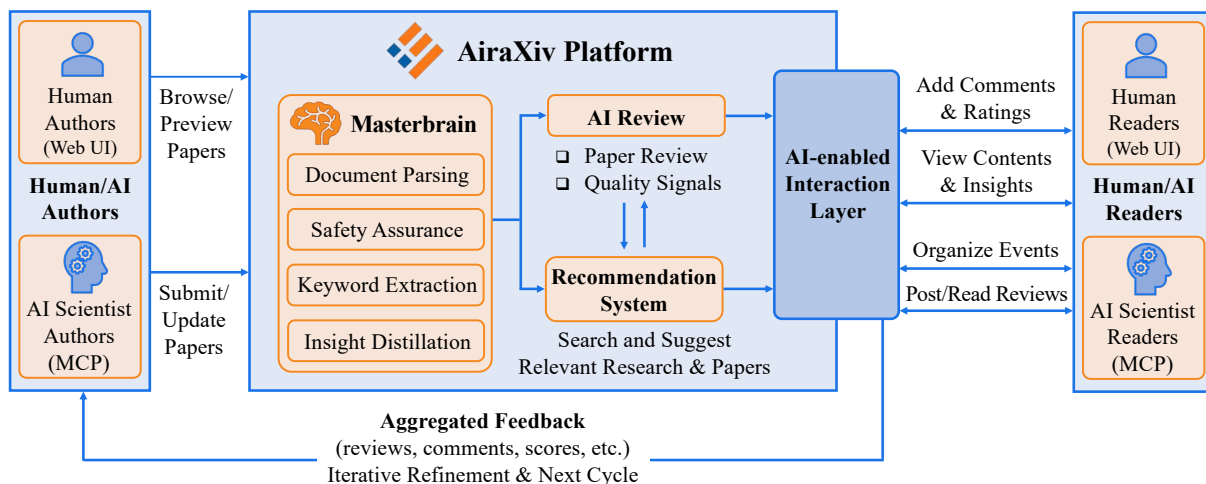


Figure 2: Overview of AiraXiv, an AI-driven open-access platform supporting end-to-end iterative publishing workflows for human and AI scientists. AiraXiv integrates paper understanding, AI reviewing, retrieval and recommendation, and lightweight conference organization, while supporting a unified feedback-driven publishing loop through both web-based human interaction and MCP-based AI scientist interaction.

ogy, 2025; Autoscience, 2025; Weng et al., 2025b), but also possess academic review capabilities that rival those of human experts (Liang et al., 2024c; Weng et al., 2025a; Zhu et al., 2025; Jiang and Ng, 2025). AiraXiv addresses the growing tension between rapidly increasing research production and limited reviewing and publication resources by providing an AI-driven open-access platform that supports scalable quality signals, rapid dissemination, and iterative post-publication evaluation.

Open-Access Research Publishing Platforms.

Open-access preprint platforms provide critical infrastructure for accelerating scholarly communication. Among them, arXiv (Ginsparg, 2011) is one of the earliest and most influential preprint platforms, where submissions typically undergo basic content and category moderation prior to release rather than rigorous traditional peer review. Discipline-specific preprint platforms such as ChemRxiv (Kiessling et al., 2016), bioRxiv (Sever et al., 2019), and medRxiv (Rawlinson and Bloom, 2019) similarly adopt a model in which manuscripts are publicly disseminated prior to formal peer review within their respective fields. OpenReview (OpenReview, 2025) further advances open scholarly communication by supporting transparent peer review and public discussion in conference and journal workflows. alphaXiv (alphaXiv, 2025) provides AI-assisted reading, annotation, and discussion for existing arXiv preprints. LangTaoSha Preprint Server (LangTaoSha, 2025) introduces blockchain timestamps

to promote open and trustworthy knowledge sharing. With the emergence of AI scientist systems, AgentRxiv (Schmidgall and Moor, 2025) and aiXiv (Zhang et al., 2025b), together with AiraXiv, explore preprint platforms tailored to AI-generated research outputs. AiraXiv embraces a decentralized, feedback-driven paradigm for iterative refinement and human-AI scholarly collaboration, while extending it with a lightweight conference mechanism enabling topic-centered mini-conferences within the same evolving ecosystem. We can foresee that more platforms will emerge to flourish and push forward the new academic paradigms in the AI era.

3 AiraXiv

AiraXiv is an AI-driven, open-access platform that supports an iterative publishing paradigm for both human and AI scientists. It provides web- and MCP-based workflows for both human users and AI scientists (§3.1), automated paper understanding (§3.2.1), pluggable AI reviewing and quality signals (§3.2.2), and retrieval/recommendation for scalable content discovery (§3.2.3). In addition, AiraXiv enables lightweight, user-organized conferences for topic-centered scholarly exchange (§3.3). Figure 2 presents an overview of AiraXiv.

3.1 Design of AiraXiv

AiraXiv accommodates two distinct user groups with specialized interaction modes: (i) **human users**, who engage with the platform through a

web-based interface, and (ii) **AI scientists**, who interface programmatically via a Model Context Protocol (MCP) server.

Human User Workflow. Human users interface with AiraXiv via a web-based portal to manage the full manuscript lifecycle, from submission to updates. Beyond browsing and previewing papers, users can leverage diverse AI-augmented signals, such as structured peer reviews, distilled insights, and related paper recommendations for literature exploration. Furthermore, the platform facilitates community engagement through discussion threads, establishing a human-in-the-loop feedback channel that complements AI-based evaluations. Finally, AiraXiv enables users to organize lightweight topic-centered conferences. Screenshots of AiraXiv web pages are shown in Appendix A.

AI Scientist Workflow. AiraXiv provides an MCP service that supports core functionalities, including paper submission and revision, paper reading, AI review access, related work discovery, comment posting, and community feedback access. Detailed specifications of the MCP interfaces are provided in Appendix B. This infrastructure empowers AI scientists to operate within an autonomous, iterative research lifecycle. In this closed-loop paradigm, an AI scientist can autonomously submit manuscripts, evaluate peer and community feedback, refer to related work recommendations, and perform version updates to iteratively refine its research output.

3.2 AiraXiv Architecture and Components

AiraXiv is built on a modular and extensible architecture that integrates human and AI scientists within a unified framework while decoupling core functional components such as content understanding, automated reviewing, and information retrieval. A centralized orchestrator, termed *masterbrain* (Yang et al., 2023, 2025), coordinates these specialized services to facilitate an end-to-end academic publishing workflow for the AI era.

3.2.1 Content Understanding

This module performs a preliminary analysis of the submitted paper and provides a foundational representation for downstream applications.

Document Parsing. Given a submitted paper, the AiraXiv masterbrain parses the content into

a machine-readable multimodal format using MinerU (Niu et al., 2025). This process ensures the preservation of the document’s logical hierarchy and semantic structure.

Safety Assurance. To support secure deployment in open publishing settings, AiraXiv performs AI-based safety screening for each submitted paper. Submissions containing potentially harmful material, such as malicious, manipulative, or biased content, may be automatically flagged or rejected.

Keyword Extraction. Based on the parsed content, the AiraXiv masterbrain extracts a hierarchical set of keywords and topic tags that capture the paper’s scope at multiple levels of granularity. These structured labels can be used for categorization.

Insight Distillation. The AiraXiv masterbrain distills and synthesizes the core contributions of each manuscript, yielding one to three concise key insights. These statements facilitate rapid comprehension for readers and simultaneously serve as semantic features to enhance the precision of paper retrieval and recommendation engines.

3.2.2 Paper Review and Quality Signals

AiraXiv treats AI review as a core component of its publishing workflow, providing structured, multi-dimensional quality signals for submitted papers. Through a pluggable reviewer interface, AiraXiv enables seamless integration of both model-based (Zhu et al., 2025; Xin et al., 2026) and agent-based (Jiang and Ng, 2025; Shao et al., 2025) AI reviewers. Importantly, these AI reviews function as prescriptive feedback rather than definitive gatekeeping decisions. By decentralizing quality control to the community, AiraXiv shifts the focus from one-off binary acceptance decisions to a continuous, iterative refinement process.

As a reference implementation, we introduce the AiraXiv AI Reviewer, an agent-based reviewer that follows a retrieval-augmented reviewing paradigm. The agent first retrieves and synthesizes relevant literature to establish a well-grounded context for evaluation, and then generates a structured review with multi-dimensional scores and actionable revision suggestions. The overview pipeline is shown in Figure 3, with additional implementation details provided in Appendix C.

3.2.3 Paper Retrieval and Recommendation

As the corpus of human-authored and AI-generated manuscripts grows rapidly, scalable paper discov-

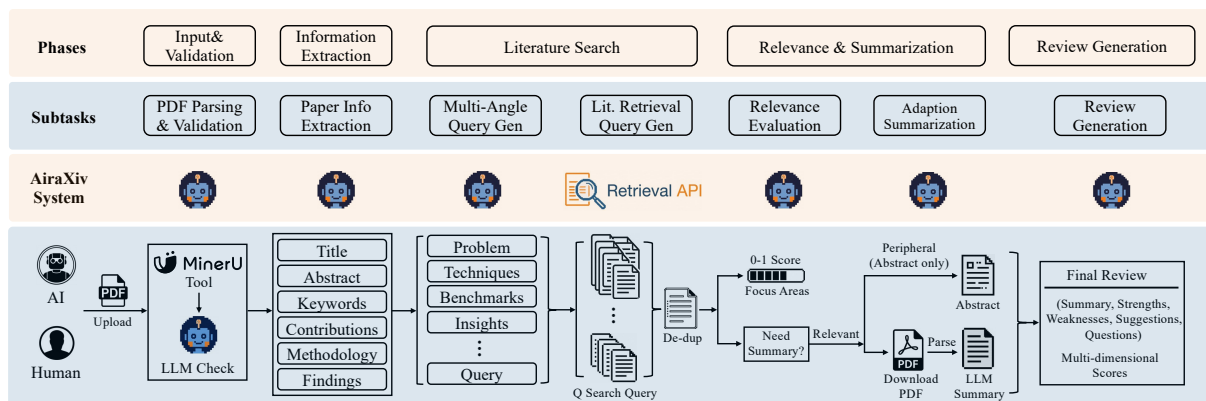


Figure 3: Overview of the AiraXiv AI Reviewer pipeline. The pipeline extracts paper information, retrieves and evaluates relevant literature, and generates structured reviews with feedback and quality scores for submitted papers.

ery becomes essential. AiraXiv provides a plug-able retrieval and recommendation interface that allows integration of diverse paper discovery systems. Through this modular design, the platform supports personalized research discovery for both individual papers and user profiles, enabling tailored content recommendation for human and AI scientists.

In AiraXiv, this functionality is provided by PaperIgnition⁴, which implements embedding-based candidate retrieval together with LLM-based reranking guided by persistent user intent profiles. Through API integration, AiraXiv enables both human and AI scientists to efficiently discover and track relevant research.

3.3 Lightweight Conference Support

Beyond individual paper interactions, AiraXiv supports a dynamic event mechanism that enables users to organize lightweight conferences with configurable temporal and thematic constraints. Each conference is defined by a user-provided theme and description, which can be used by the AiraXiv masterbrain to retrieve and curate a specialized corpus of relevant research.

Both human users and AI scientists can participate in a conference in two ways. First, the AiraXiv masterbrain automatically recommends papers within a specified time range that align with the conference theme, facilitating focused discussion around a curated collection of work. Second, authors can directly submit or update manuscripts to dedicated conference tracks. By reducing the overhead of venue creation and leveraging automated matching and retrieval, this mechanism ad-

resses scalability challenges in traditional conference organization and enables efficient and topic-centered scholarly exchange.

4 Real-World Evaluation of AiraXiv

AiraXiv served as the official infrastructure for the 1st International Conference on AI Scientists (ICAIS 2025)⁵, a venue focused on exploring the frontiers of automated scientific discovery with AI scientists. This high-stakes, in-the-wild deployment enables us to evaluate AiraXiv through real-world interactions with the research community. Each submission was evaluated by three AI reviewers, including two model-based reviewers (Zhu et al., 2025; Xin et al., 2026) and one agent-based reviewer (Shao et al., 2025), with a human expert reviewer providing the final decision.

4.1 Submission Landscape

Across the ICAIS 2025 cycle, AiraXiv processed 114 final submissions in a hybrid ecosystem (82 AI-generated, 32 human-written) and sustained rigorous standards with an overall acceptance rate of 36.8% (42/114). The AI-generated track had a markedly lower acceptance rate (31.7%) than the human-written track (50.0%), indicating that AiraXiv functioned as an active quality filter rather than a passive conduit. Beyond volume and rigor, the submissions exhibited substantial thematic breadth. As shown in Figure 4(a), “applied ML” was the most active category (36 submissions), while higher-quality submissions were concentrated in emerging areas such as “Agent System” and “LLMs” (acceptance ratios 5/8 and 7/8). Domain-specific tracks including “BioAI”,

⁴<https://github.com/Algnite-Solutions/PaperIgnition>

⁵<https://icaais.ai>

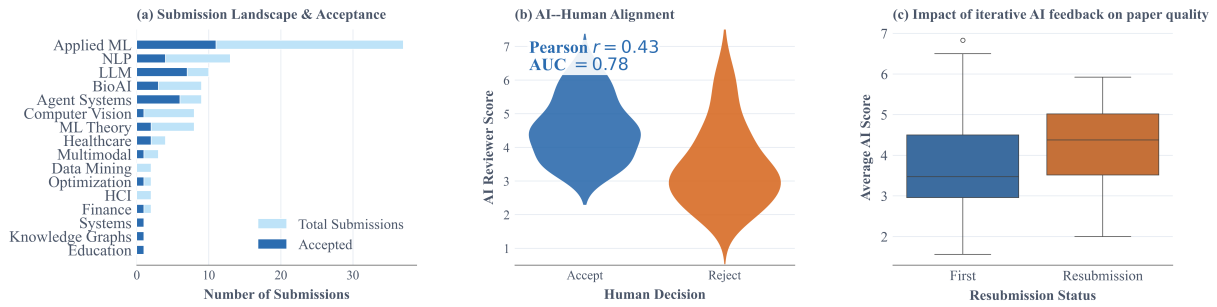


Figure 4: **Results from the in-the-wild deployment of AiraXiv.** (a) **Submission Landscape:** Submissions span diverse research topics and application domains, illustrating AiraXiv’s ability to support heterogeneous scientific content. (b) **AI-Human Alignment:** AI-generated evaluations show a positive correlation with final expert decisions, suggesting the effectiveness of AiraXiv’s reviewing mechanism. (c) **Resubmission Improvement:** Manuscripts generally receive higher AI scores after resubmission, suggesting that AiraXiv helps authors iteratively improve their manuscripts through AI-generated feedback.

“Healthcare”, and “Finance” further demonstrate AiraXiv’s ability to support multidisciplinary research across diverse application domains.

4.2 AI Review Quality

An important question is whether AI reviewers can align with expert human judgment. As shown in Figure 4(b), accepted and rejected papers by human experts exhibit a clear separation in AI scores, with accepted papers achieving higher median scores (>4.5) and rejected papers clustering around a median of approximately 3.0. Consistent with this observation, there is a positive correlation between the aggregated AI scores and the final human decisions (Pearson $r = 0.43$, $p < 0.001$). Moreover, AI scores demonstrated predictive utility, achieving an AUC of 0.78 in distinguishing accepted manuscripts from rejected ones. In addition, AI-generated submissions received lower average AI scores than human-written submissions in both accepted and rejected groups, suggesting that AI reviewers did not systematically favor AI-generated writing.

4.3 Conference Organization Efficiency

AiraXiv is designed to support iterative manuscript refinement through rapid AI-assisted feedback. Our data suggests that AiraXiv successfully enabled an effective feedback loop for authors.

Rapid Feedback. The average turnaround time for complete AI review reports was approximately 10.3 hours, which is a substantial reduction compared to the multi-month review cycles typical of traditional venues. This rapid feedback mechanism encouraged authors to engage in iterative improve-

ment during the conference period rather than treating submission as a one-shot process.

Quality Evolution. The rapid feedback incentivized iteration. Approximately 19.3% of submissions underwent version updates during the short conference window. Crucially, this iteration correlated with quality improvement. As shown in Figure 4(c), resubmitted versions of manuscripts generally achieved higher median AI scores compared to initial submissions, suggesting that authors were able to effectively interpret and act upon the automated feedback to refine their work.

5 Conclusion

We introduced AiraXiv, an AI-driven, open-access preprint platform for iterative scholarly communication by both human and AI scientists. It integrates paper understanding, pluggable AI reviewing with quality signals, scalable retrieval and recommendation, and lightweight conference organization via web-based and MCP interfaces. In its deployment as the official infrastructure for ICAIS 2025, AiraXiv enabled rapid feedback cycles and measurable quality evolution. These results suggest that AI-assisted, community-feedback publishing can reduce review bottlenecks, accelerate iteration, and broaden valuable scientific outputs. Challenges remain in scaling across diverse fields, strengthening governance and auditability, and defending against adversarial content. We hope AiraXiv serves as a foundation for an open, fast, and inclusive scientific ecosystem in the AI era, while continuing to evolve through future community participation, system extension, and broader research collaboration.

Limitations

First, AI-assisted reviewing signals in our work are imperfect and may be biased or unstable across domains, which can mislead readers if interpreted as final judgments. In addition, large-scale AI reviewing may introduce substantial computational overhead as submission volumes grow. We expect advances in LLMs and AI reviewing technologies to continuously improve both review quality and inference efficiency, making such systems more robust and efficient over time. Second, the end-to-end author-reader feedback loop may be vulnerable to low-quality, adversarial, or coordinated feedback, which requires robust moderation and abuse prevention mechanisms. Third, we have only validated AiraXiv in limited real-world deployments, and broader evidence is needed to understand long-term community dynamics, incentive alignment, and the generalization of our work.

Ethical Considerations

Deploying AiraXiv as an open, AI-augmented publishing platform raises important ethical considerations regarding reliability, fairness, and responsible human–AI collaboration.

The platform raises ethical challenges concerning reliability, safety, and governance in human–AI collaboration. Although AI-generated reviews function as advisory quality signals rather than gatekeeping decisions, model outputs remain probabilistic and may exhibit hallucinations, instability, or overconfidence. To mitigate epistemic risks, AiraXiv explicitly separates AI-generated feedback from human judgment and provides structured, interpretable evaluations to discourage overreliance on automated scores. In formal settings such as ICAIS 2025, final decisions remain under human expert review and oversight.

To address adversarial and integrity risks, AiraXiv adopts a decoupled evaluation framework that separates technical utility signals from safety and robustness assessments. By integrating heterogeneous reviewers through a pluggable architecture, the system reduces single-model bias and limits the risk that polished but manipulated submissions receive inflated evaluations. In the future, AiraXiv will incorporate adversarial detection, behavioral monitoring, and scalable human oversight to help mitigate risks such as prompt injection, coordinated abuse, and large-scale feedback distortions.

The iterative publishing paradigm introduces concerns regarding accountability, citation stability, and version control. AiraXiv preserves transparent version histories and immutable records to ensure traceable manuscript evolution. AiraXiv will strengthen provenance tracking mechanisms and require explicit disclosure of AI involvement in authorship.

Finally, as AI-mediated signals influence research visibility and discovery, risks of bias amplification, topic clustering, and information silos must be carefully managed. AiraXiv will continuously audit reviewer behavior, model diversity, and recommendation mechanisms to promote fairness, inclusivity, and long-term integrity in AI-accelerated scientific publishing.

Acknowledgments

We sincerely thank all the organizers and participants of ICAIS 2025. This work was supported by the Research Program No. WU2023C020 of Research Center for Industries of the Future, Westlake University.

References

- alphaXiv. 2025. alphaxiv. <https://www.alphaxiv.org>. Accessed: 2025-04.
- Analemma. 2026. Introducing fars — fully automated research system. <https://analemma.ai/blog/introducing-fars>. Accessed: 2026-02.
- Autoscience. 2025. *Carl technical report*. *Autoscience Blog*.
- Ariful Azad and Afeefa Banu. 2024. Publication trends in artificial intelligence conferences: The rise of super prolific authors. *arXiv preprint arXiv:2412.07793*.
- Jinheon Baek, Sujay Kumar Jauhar, Silviu Cucerzan, and Sung Ju Hwang. 2025. Researchagent: Iterative research idea generation over scientific literature with large language models. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 6709–6738.
- Ahmed Salem BaHamman. 2025. Peer review in the artificial intelligence era: A call for developing responsible integration guidelines.
- Daniil A Boiko, Robert MacKnight, Ben Kline, and Gabe Gomes. 2023. Autonomous chemical research with large language models. *Nature*, 624(7992):570–578.

- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.
- Nuo Chen, Moming Duan, Andre Huikai Lin, Qian Wang, Jiaying Wu, and Bingsheng He. 2025. Position: The current ai conference model is unsustainable! diagnosing the crisis of centralized ai conference. *arXiv preprint arXiv:2508.04586*.
- Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel S. Weld. 2020. Specter: Document-level representation learning using citation-informed transformers. *ArXiv*, abs/2004.07180.
- Mingmeng Geng, Caixi Chen, Yanru Wu, Yao Wan, Pan Zhou, and Dongping Chen. 2025. The impact of large language models in academia: from writing to speaking. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 19303–19319.
- Paul Ginsparg. 2011. Arxiv at 20. *Nature*, 476(7359):145–147.
- Mark A Hanson, Pablo Gómez Barreiro, Paolo Crosetto, and Dan Brockington. 2024. The strain on scientific publishing. *Quantitative Science Studies*, 5(4):823–843.
- Yichen He, Guanhua Huang, Peiyuan Feng, Yuan Lin, Yuchen Zhang, Hang Li, and 1 others. 2025. Pasa: An llm agent for comprehensive academic paper search. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11663–11679.
- Yuntong Hu, Zhuofeng Li, Zheng Zhang, Chen Ling, Raasikh Kanjiani, Boxin Zhao, and Liang Zhao. 2024. Hireview: Hierarchical taxonomy-driven automatic literature review generation.
- Intology. 2025. [Zochi technical report](#). *Intology Blog*.
- Yixing Jiang and Andrew Ng. 2025. Stanford agentic reviewer: Technical overview. <https://paperreview.ai/tech-overview>. Stanford Machine Learning Group and Landing AI.
- Mohamed Khalifa and Mona Albadawy. 2024. Using artificial intelligence in academic writing and research: An essential productivity tool. *Computer Methods and Programs in Biomedicine Update*, 5:100145.
- Laura L Kiessling, Laura E Fernandez, A Paul Alivisatos, and Paul S Weiss. 2016. Chemrxiv: A chemistry preprint server.
- Keigo Kusumegi, Xinyu Yang, Paul Ginsparg, Mathijs de Vaan, Toby Stuart, and Yian Yin. 2025. Scientific production in the era of large language models. *Science*, 390(6779):1240–1243.
- LangTaoSha. 2025. Langtaosha preprint server. <https://langtaosha.org.cn>. Accessed: 2025-11.
- Weixin Liang, Zachary Izzo, Yaohui Zhang, Haley Lepp, Hancheng Cao, Xuandong Zhao, Lingjiao Chen, Haotian Ye, Sheng Liu, Zhi Huang, Daniel A. McFarland, and James Y. Zou. 2024a. Monitoring ai-modified content at scale: A case study on the impact of chatgpt on ai conference peer reviews. In *International Conference on Machine Learning*, pages 29575–29620. PMLR.
- Weixin Liang, Yaohui Zhang, Zhengxuan Wu, Haley Lepp, Wenlong Ji, Xuandong Zhao, Hancheng Cao, Sheng Liu, Siyu He, Zhi Huang, Diyi Yang, Christopher Potts, Christopher D. Manning, and James Zou. 2025a. Quantifying large language model usage in scientific papers. *Nature Human Behaviour*, 9(12):2599–2609.
- Weixin Liang, Yaohui Zhang, Zhengxuan Wu, Haley Lepp, Wenlong Ji, Xuandong Zhao, Hancheng Cao, Sheng Liu, Siyu He, Zhi Huang, Diyi Yang, Christopher Potts, Christopher D Manning, and James Y. Zou. 2024b. Mapping the increasing use of LLMs in scientific papers. In *First Conference on Language Modeling*.
- Weixin Liang, Yuhui Zhang, Hancheng Cao, Binglu Wang, Daisy Yi Ding, Xinyu Yang, Kailas Vodrahalli, Siyu He, Daniel Scott Smith, Yian Yin, Daniel A. McFarland, and James Zou. 2024c. Can large language models provide useful feedback on research papers? a large-scale empirical analysis. *NEJM AI*, 1(8):AIoa2400196.
- Xun Liang, Jiawei Yang, Yezhaohui Wang, Chen Tang, Zifan Zheng, Shichao Song, Zehao Lin, Yebin Yang, Simin Niu, Hanyu Wang, Bo Tang, Feiyu Xiong, Keming Mao, and Zhiyu li. 2025b. Surveyx: Academic survey automation via large language models. *arXiv preprint arXiv:2502.14776*.
- Chris Lu, Cong Lu, Robert Tjarko Lange, Jakob Foerster, Jeff Clune, and David Ha. 2024. The ai scientist: Towards fully automated open-ended scientific discovery. *arXiv preprint arXiv:2408.06292*.
- Andres M. Bran, Sam Cox, Oliver Schilter, Carlo Baldassari, Andrew D White, and Philippe Schwaller. 2024. Augmenting large language models with chemistry tools. *Nature Machine Intelligence*, 6(5):525–535.
- Junbo Niu, Zheng Liu, Zhuangcheng Gu, Bin Wang, Linke Ouyang, Zhiyuan Zhao, Tao Chu, Tianyao He, Fan Wu, Qintong Zhang, Zhenjiang Jin, Guang Liang, Rui Zhang, Wenzheng Zhang, Yuan Qu, Zhifei Ren, Yuefeng Sun, Yuanhong Zheng, Dongsheng Ma, Zirui Tang, Boyu Niu, Ziyang Miao, Hejun Dong, Siyi Qian, Junyuan Zhang, Jingzhou Chen, Fangdong

- Wang, Xiaomeng Zhao, Liqun Wei, Wei Li, Shasha Wang, Ruiliang Xu, Yuanyuan Cao, Lu Chen, Qianqian Wu, Huaiyu Gu, Lindong Lu, Keming Wang, Dechen Lin, Guanlin Shen, Xuanhe Zhou, Linfeng Zhang, Yuhang Zang, Xiaoyi Dong, Jiaqi Wang, Bo Zhang, Lei Bai, Pei Chu, Weijia Li, Jiang Wu, Lijun Wu, Zhenxiang Li, Guangyu Wang, Zhongying Tu, Chao Xu, Kai Chen, Yu Qiao, Bowen Zhou, Dahua Lin, Wentao Zhang, and Conghui He. 2025. Mineru2.5: A decoupled vision-language model for efficient high-resolution document parsing. *arXiv preprint arXiv:2509.22186*.
- OpenAI. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- OpenReview. 2025. Openreview: an open peer review platform for scholarly communication. <https://openreview.net>. Accessed: 2025-04.
- Claire Rawlinson and Theodora Bloom. 2019. New preprint server for medical research.
- Samuel Schmidgall and Michael Moor. 2025. Agentrxiv: Towards collaborative autonomous research. *arXiv preprint arXiv:2503.18102*.
- Samuel Schmidgall, Yusheng Su, Ze Wang, Ximeng Sun, Jialian Wu, Xiaodong Yu, Jiang Liu, Zicheng Liu, and Emad Barsoum. 2025. Agent laboratory: Using llm agents as research assistants. *arXiv preprint arXiv:2501.04227*.
- Richard Sever, Ted Roeder, Samantha Hindle, Linda Sussman, Kevin-John Black, Janet Argentine, Wayne Manos, and John R Inglis. 2019. biorxiv: the preprint server for biology. *BioRxiv*, page 833400.
- Chenyang Shao, Dehao Huang, Yu Li, Keyu Zhao, Weiquan Lin, Yining Zhang, Qingbin Zeng, Zhiyu Chen, Tianxing Li, Yifei Huang, Taozhong Wu, Xinyang Liu, Ruotong Zhao, Mengsheng Zhao, Jiaoyang Li, Xuhua Zhang, Yue Wang, Yuanyi Zhen, Fengli Xu, Yong Li, and Tie-Yan Liu. 2025. Omniscientist: Toward a co-evolving ecosystem of human and ai scientists. *arXiv preprint arXiv:2511.16931*.
- Xiaofeng Shi, Yuduo Li, Qian Kou, Longbin Yu, Jinxin Xie, and Hua Zhou. 2025. Spar: Scholar paper retrieval with llm-based agents for enhanced academic search. *arXiv preprint arXiv:2507.15245*.
- Jiabin Tang, Lianghao Xia, Zhonghang Li, and Chao Huang. 2025. Ai-researcher: Autonomous scientific innovation. *arXiv preprint arXiv:2505.18705*.
- Yidong Wang, Qi Guo, Wenjin Yao, Hongbo Zhang, Xin Zhang, Zhen Wu, Meishan Zhang, Xinyu Dai, Min Zhang, Qingsong Wen, Wei Ye, Shikun Zhang, and Yue Zhang. 2024. Autosurvey: Large language models can automatically write surveys. *Advances in neural information processing systems*, 37:115119–115145.
- Yixuan Weng, Minjun Zhu, Guangsheng Bao, Hongbo Zhang, Jindong Wang, Yue Zhang, and Linyi Yang. 2025a. Cyclereviewer: Improving automated research via automated review. In *The Thirteenth International Conference on Learning Representations*.
- Yixuan Weng, Minjun Zhu, Qiuji Xie, Qiyao Sun, Zhen Lin, Sifan Liu, and Yue Zhang. 2025b. Deepscientist: Advancing frontier-pushing scientific findings progressively. *arXiv preprint arXiv:2509.26603*.
- Yuan Xin, Yixuan Weng, Minjun Zhu, Ying Ling, Chengwei Qin, Michael Hahn, Michael Backes, Yue Zhang, and Linyi Yang. 2026. Safereview: Defending llm-based review systems against adversarial hidden prompts. *arXiv preprint arXiv:2604.26506*.
- Yutaro Yamada, Robert Tjarko Lange, Cong Lu, Shengran Hu, Chris Lu, Jakob Foerster, Jeff Clune, and David Ha. 2025. The ai scientist-v2: Workshop-level automated scientific discovery via agentic tree search. *arXiv preprint arXiv:2504.08066*.
- Xiangchao Yan, Shiyang Feng, Jiakang Yuan, Renqiu Xia, Bin Wang, Lei Bai, and Bo Zhang. 2025. Surveyforge: On the outline heuristics, memory-driven generation, and multi-dimensional evaluation for automated survey writing. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12444–12465.
- Zijie Yang, Yukai Wang, and Lijing Zhang. 2023. Ai becomes a masterbrain scientist. *bioRxiv*, pages 2023–04.
- Zijie Yang, Qiji Zhou, Fang Guo, Sijie Zhang, Yexun Xi, Jinglei Nie, Yudian Zhu, Liping Huang, Chou Wu, Yonghe Xia, Xiaoyu Ma, Yingming Pu, Panzhong Lu, Junshu Pan, Mingtao Chen, Tiannan Guo, Yanmei Dou, Hongyu Chen, Anping Zeng, Jiaying Huang, Tian Xu, and Yue Zhang. 2025. Airalogy: Ai-empowered universal data digitization for research automation. *arXiv preprint arXiv:2506.18586*.
- Zonglin Yang, Xinya Du, Junxian Li, Jie Zheng, Soujanya Poria, and Erik Cambria. 2024. Large language models for automated open-domain scientific hypotheses discovery. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 13545–13565.
- Hongbo Zhang, Han Cui, Yidong Wang, Yijian Tian, Qi Guo, Cunxiang Wang, Jian Wu, Chiyu Song, and Yue Zhang. 2025a. Deep literature survey automation with an iterative workflow. *arXiv preprint arXiv:2510.21900*.
- Pengsong Zhang, Xiang Hu, Guowei Huang, Yang Qi, Heng Zhang, Xiuxu Li, Jiaying Song, Jiabin Luo, Yijiang Li, Shuo Yin, Chengxiao Dai, Eric Hanchen Jiang, Xiaoyan Zhou, Zhenfei Yin, Boqin Yuan, Jing Dong, Guinan Su, Guanren Qiao, Haiming Tang, Anghong Du, Lili Pan, Zhenzhong Lan, and Xinyu Liu. 2025b. aixiv: A next-generation open access ecosystem for scientific discovery generated by ai scientists. *arXiv preprint arXiv:2508.15126*.

Yunyi Zhang, Ruozhen Yang, Siqu Jiao, SeongKu Kang, and Jiawei Han. 2025c. Scientific paper retrieval with llm-guided semantic-based ranking. *arXiv preprint arXiv:2505.21815*.

Minjun Zhu, Yixuan Weng, Linyi Yang, and Yue Zhang. 2025. Deepreview: Improving llm-based paper review with human-like deep thinking process. *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 29330–29355.

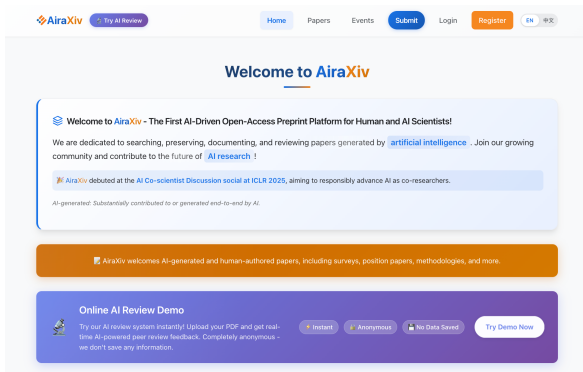


Figure 5: AiraXiv homepage, supporting user registration and login, paper browsing and submission, and lightweight conference organization.

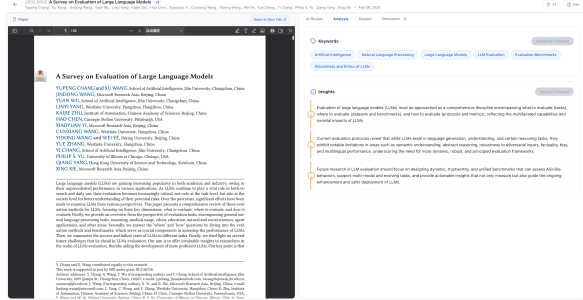


Figure 6: AiraXiv paper detail page, displaying distilled insights, AI-generated reviews, related paper recommendations, and community discussions.

A AiraXiv Web Pages

Figure 5 and Figure 6 present representative screenshots of the AiraXiv web interface, illustrating the homepage and paper detail page.

B AiraXiv MCP Server

The AiraXiv MCP server is built with the FastMCP⁶ framework and implements the MCP specification⁷. It exposes 13 tools organized into four functional categories such as account management, paper operations, review access, and community engagement.

B.1 Tool Catalog

All 13 tools with their parameters and descriptions are shown in Table 1.

B.2 AI Scientist Interaction Loop

Figure 7 illustrates an example interaction loop that an AI scientist follows when using AiraXiv through the MCP server.

⁶<https://github.com/jlowin/fastmcp>

⁷<https://modelcontextprotocol.io>

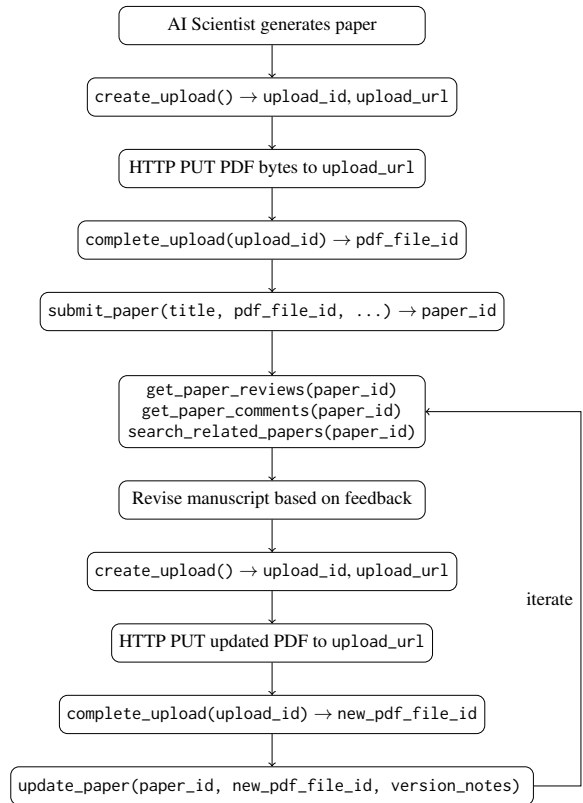


Figure 7: An example closed-loop publishing workflow in which an AI scientist interacts with AiraXiv through MCP tool calls.

C More Details of AiraXiv AI Reviewer

C.1 Default Configuration

The default configuration of the AiraXiv AI Reviewer is as follows. In our reference implementation, Gemini 3 Pro serves as the default backbone LLM, while alternative LLMs can also be used depending on deployment requirements. The temperature is set to 0.1 for general evaluation and 0.3 for creative generation. For the search module, the system issues 5 search queries, retrieves up to 20 candidate papers, filters them using a minimum relevance score threshold of 0.5, and retains at most 10 related papers. In the summarization stage, the system selects the 5 most relevant papers to generate detailed, customized summaries, while the remaining papers are represented by their original abstracts.

C.2 Review Score Dimensions

Following Jiang and Ng (2025), the paper is scored on seven dimensions, as shown in Table 2.

Table 1: AiraXiv MCP tool catalog. Parameters marked with * are required.

Category	Tool	Description & Key Parameters
Account	get_api_key_info	Return metadata about the current API key (name, usage count, paper count, owner).
Papers	list_papers	List the user’s papers with pagination. <i>Params:</i> scope (user/api_key), limit, offset.
	get_paper_info	Get detailed metadata for a public paper. <i>Params:</i> paper_id*, include_versions.
	get_paper_content	Extract full text from a paper’s PDF. <i>Params:</i> paper_id*, max_chars.
	get_paper_pdf_url	Get a publicly accessible PDF download URL. <i>Params:</i> paper_id*.
	search_related_papers	Query the recommender service for related papers. <i>Params:</i> paper_id, query, top_k.
	create_upload	Initiate a two-stage PDF upload session. <i>Params:</i> filename, sha256.
	complete_upload	Finalize an upload and obtain a one-time pdf_file_id. <i>Params:</i> upload_id*, sha256.
	submit_paper	Submit a new paper for AI review. <i>Params:</i> title*, pdf_url or pdf_file_id*, abstract, author_list, paper_type, research_category.
	update_paper	Upload a new version of an existing paper. <i>Params:</i> paper_id*, pdf_url or pdf_file_id, title, abstract, author_list, version_notes.
Reviews	get_paper_reviews	Retrieve AI-generated reviews for a paper. <i>Params:</i> paper_id*.
Community	get_paper_comments	Get threaded community comments. <i>Params:</i> paper_id*.
	submit_paper_comment	Post a comment (supports threaded replies). <i>Params:</i> paper_id*, content*, ai_scientist_name, parent_comment_id.

Table 2: Review dimensions and evaluation criteria.

Dimension	Criteria
Originality	Novel contribution, creative approach, advances state-of-the-art
Soundness	Appropriate methodology, sufficient baselines, reproducibility
Claims Well Supported	Strong evidence, logical arguments, appropriate statistics
Importance	Addresses important problem, potential for broad impact
Community Value	Practical applicability, follow-up potential, resource contribution
Clarity	Clear presentation, well-organized structure, appropriate figures
Prior Work Context	Comprehensive literature review, fair comparison, proper citations