

A Preprint Republic at a Crossroads: Governance, Transparency, and the Future of arXiv as a Public Good

Renato Vieira dos Santos*

*Instituto de Ciência, Tecnologia e Inovação – ICTIN,
Universidade Federal de Lavras – UFLA, Campus Paraíso, MG 37950-000, Brazil*

The arXiv preprint repository, founded in 1991, has become the de facto dissemination platform for physics, mathematics, computer science, and increasingly other quantitative disciplines, hosting over 2.4 million articles and serving millions of users worldwide. In mid-2026, its governance will transition from Cornell University Library to an independent nonprofit entity governed by a board of directors and supported by private philanthropy. This article provides a data-driven examination of that transition, placing it within the history of the repository and the broader political economy of scholarly communication. Drawing on public submission statistics, moderation records, documented controversies, and historical precedents from academic publishing, we identify six structural risks: opaque governance, centralized editorial control, financial unsustainability, systemic bias, erosion of scientific pluralism, and potential mission drift toward privatized gatekeeping. We argue that the moment before institutional inertia sets in is precisely when the research community must insist on binding transparency and accountability mechanisms. We propose a set of concrete safeguards—a community-elected advisory council, mandatory transparency reporting, an independent moderation ombuds, a free-access charter, and periodic external audits—that can preserve arXiv’s identity as a public good without compromising operational agility. The manuscript itself has been submitted to arXiv as a real-time test of the moderation system’s openness to self-examination.

Keywords: arXiv; scholarly communication; open access; governance; preprint repositories; scientific infrastructure; generative artificial intelligence.

I. INTRODUCTION

In the *Republic*, Plato entrusted the guardians of the city with a paradoxical duty: to possess knowledge of the highest order while remaining transparent to the citizens they served. The tension between epistemic authority and democratic accountability has never been resolved; it has only migrated from the polis to the digital infrastructures that now govern the circulation of scientific knowledge. Among these, no platform is more consequential than arXiv.

Conceived by Paul Ginsparg at Los Alamos National Laboratory in August 1991, arXiv began as a modest automated email server that allowed a few hundred high-energy physicists to share preprints [1, 2]. Within three decades it has expanded to host more than 2.4 million articles, with monthly submission rates surpassing 17,000 in 2024 [3]. The repository is accessed by millions of distinct users each month, and its content generates hundreds of millions of full-text downloads annually, making it among the most heavily used single-source scientific literature platforms globally [4]. In disciplines such as high-energy physics, cosmology, and parts of mathematics, the arXiv preprint has effectively displaced the journal article as the primary vehicle for rapid dissemination of results.

The distinctive feature of arXiv is its radical openness: any researcher with a valid institutional affiliation can deposit a paper, which becomes freely available worldwide

within hours, without prior peer review. This zero-cost, zero-embargo model has made arXiv a cornerstone of the open-access movement and a counterweight to the serials crisis that has concentrated academic publishing in the hands of a few large commercial houses [5, 6]. At a time when average article-processing charges for gold open-access journals exceed two thousand dollars, arXiv remains free at the point of use. It is, therefore, not merely a database but a piece of public-interest infrastructure, comparable in function to a public library or a publicly funded research network.

From its inception until mid-2026, arXiv was hosted and operated by the Cornell University Library, with financial support from the Simons Foundation[34], member institutions, and other donors. This arrangement provided a recognizable governance model: ultimate responsibility rested with a single academic institution, and the service operated under the umbrella of a university library whose mission aligns with preservation and open access. In early 2025, it was announced that, effective July 2026, arXiv would separate from Cornell and become a fully independent nonprofit organization, governed by its own board of directors and led by a newly appointed CEO [7]. The Simons Foundation has committed a substantial multi-million dollar endowment to anchor the new entity, and a board of directors has been formed, drawing members from academia, philanthropy, and industry.

This transition raises a set of questions that, we believe, the global user community has not yet examined with sufficient rigor. Independence from a single institution may bring agility, but it also removes a layer of insti-

* renato.santos@ufla.br

tutional accountability that, however imperfect, was legible and contestable. The board’s composition, its mechanisms for soliciting community input, the criteria for appointing moderators, and the long-term financial model are all being defined in a context that is, at present, remarkably opaque. History offers numerous cautionary tales in which scientific infrastructure, once separated from its original public-interest moorings, evolved into a private gatekeeper, with consequences that were unintended but nonetheless real [8, 9].

This paper does not accuse any individual or organization of bad faith. It adopts the stance of *structural precaution*: the systematic identification of risks that accompany governance transitions in shared cognitive infrastructure, even when all parties act with integrity. We begin by reviewing the quantitative dimensions of arXiv’s role in the research ecosystem, so that the stakes of governance are precisely understood. We then describe the new governance architecture and compare it with the preceding custodial model. The core of the paper identifies six axes of structural risk—transparency, editorial independence, financial sustainability, systemic bias, scientific pluralism, and mission drift—each illustrated with evidence from public records or documented historical episodes. For each axis, we articulate open questions that the new board should address publicly before the institutional structure crystallizes. We conclude with a set of actionable recommendations designed to embed accountability mechanisms into the new entity without undermining its operational efficiency. The manuscript itself has been submitted to arXiv as a modest empirical test of the system’s moderation processes.

II. THE ARXIV ECOSYSTEM: SCALE, IMPACT, AND STRUCTURAL SIGNIFICANCE

To appreciate the magnitude of the governance challenge, it is necessary to grasp the quantitative dimensions of arXiv’s role in the production and circulation of scientific knowledge. Table I presents a summary of submission statistics compiled from the repository’s own public records and from the scholarly literature.

The data show a sustained, roughly exponential growth that has only recently begun to level off, reflecting the saturation of the core physics disciplines and the increasing adoption of alternative preprint servers in other fields. The fraction of physics papers has declined from a near-total monopoly in the early 1990s to approximately half of submissions today, while computer science and mathematics have grown steadily, and new categories such as quantitative biology and quantitative finance have been added.

Readership data are equally telling. Total full-text downloads across all mirrors exceeded 300 million in 2023 [3], and the user base is geographically distributed, with significant traffic originating from East Asia, South Asia, and Latin America, well beyond the traditional North

American and European core. In Brazil alone, institutional downloads from arXiv rank among the top ten globally, underscoring the repository’s role as a leveling agent in regions where access to expensive journal subscriptions remains precarious.

Beyond raw volume, arXiv has become structurally embedded in the evaluation practices of the disciplines it serves. A longitudinal study by Larivière *et al.* [10] demonstrated that arXiv preprints receive a citation premium, accelerating the diffusion of results compared to non-archived articles. This effect is especially pronounced in high-energy physics, where the preprint culture effectively replaced journal-based timeliness decades ago. More broadly, the availability of preprints has been linked to increased collaboration, interdisciplinary fertilization, and faster correction of errors [11, 12]. During the COVID-19 pandemic, the role of preprint servers in accelerating the dissemination of urgent biomedical research was widely documented [13, 14], and arXiv’s existing infrastructure served as a model for the rapid deployment of related services such as medRxiv.

It is precisely because arXiv is so deeply woven into the fabric of daily scientific practice that any alteration in its governance carries systemic risk. A repository of this scale is not a mere publication venue; it is a piece of shared cognitive infrastructure whose accessibility, neutrality, and editorial integrity have direct consequences for the productivity and fairness of entire disciplines.

III. GOVERNANCE EVOLUTION: FROM CUSTODIAL OVERSIGHT TO INDEPENDENT NONPROFIT

The governance model that operated from 2001, when arXiv moved from Los Alamos to Cornell, until the announced transition, can be characterized as custodial. The Cornell University Library assumed legal and fiscal responsibility, while scientific oversight was delegated to a loose federation of subject-specific advisory committees composed of active researchers. Moderation decisions were made by volunteer scientists, typically chosen for their recognized standing in a subfield. This distributed structure, while not free of problems, diffused authority and minimized the risk of capture by any single interest group. It also maintained a direct, if informal, line of accountability to the university’s academic governance structures.

In early 2025, the leadership of arXiv and Cornell announced the creation of a new independent entity, arXiv Inc., a 501(c)(3) nonprofit organization governed by a board of directors. The transition is scheduled for completion in July 2026 [7]. The publicly stated motivation is to allow arXiv to pursue a more diversified funding base and to professionalize its operations beyond what a single university library can offer. The Simons Foundation has provided a multi-million dollar endowment to anchor the new organization, and a small number of other

TABLE I. arXiv submission volume and cumulative articles for selected years. Data compiled from Refs. [2–4].

Year	Monthly submissions	Cumulative articles	Physics fraction
1991 (Aug)	~400	400	100%
1995	~2,000	50,000	80%
2005	~4,500	350,000	67%
2015	~9,000	1,100,000	58%
2020	~14,000	1,800,000	55%
2024	~17,500	2,400,000	50%

philanthropic and institutional partners have joined as founding supporters.

As of this writing, the governance documents that will define the board’s composition, appointment procedures, and community oversight mechanisms have not been made publicly available. Press releases identify a chairperson and a few members with strong ties to elite U.S. research universities and to the Simons Foundation, but the absence of published bylaws means that the community cannot assess whether the board’s composition will reflect the geographic and disciplinary diversity of the user base, nor whether term limits or election mechanisms exist to prevent self-perpetuation [30]. Equally absent from the public record is any detailed statement about how moderators will be appointed under the new regime, whether the existing advisory committees will retain their role or be replaced, and what mechanisms, if any, will exist for community recourse against moderation decisions.

This opacity is the origin of the structural concerns elaborated below. It is important to emphasize that opacity is not, in itself, evidence of malfeasance. However, the scholarly literature on institutional governance is clear: transparency at the moment of structural creation is a strong predictor of long-term accountability [15, 16]. Once a board begins to operate under closed rules, path dependence makes subsequent opening substantially more difficult. The present moment, before institutional inertia sets in, is therefore the window during which binding transparency commitments can be secured at minimal cost.

IV. STRUCTURAL RISKS AND OPEN QUESTIONS

A. Transparency and Accountability

The new board will control a resource that directly affects the careers and daily workflows of hundreds of thousands of researchers across dozens of disciplines. Yet the global user community currently has no formal channel to scrutinize board decisions, no guarantee that board meetings will produce public minutes, and no assurance that the board’s composition will evolve to reflect the geographic and disciplinary diversity of the user base. By contrast, repositories such as Zenodo, PubMed Cen-

tral, and Open Science Framework operate with governance models that are documented in publicly available charters and are subject to regular reporting obligations [9, 29].

The arXiv transition risks replicating a pattern observed in other nonprofit conversions: a founding board is assembled from a small circle of trusted individuals, typically those who facilitated the transition, and that board subsequently becomes self-perpetuating unless explicit, legally binding term limits and election mechanisms are built into the bylaws from the outset. Without such provisions, the institution that emerges may be formally independent of any single university but effectively controlled by a philanthropic-patron model that is no more accountable to the global community than the previous arrangement.

We therefore pose the following open questions: (i) Will the bylaws of arXiv Inc. be published in full? (ii) Will board members be required to disclose conflicts of interest, including ties to commercial publishers? (iii) Will there be an annual public report that includes moderation statistics, rejection rates by subfield, and, to the extent possible under privacy constraints, the demographic distribution of accepted versus rejected authors? At present, none of these commitments has been made.

B. Editorial Independence and the Risk of Centralized Curation

arXiv’s moderation system has always existed in a delicate tension between quality control and censorship. The official policy permits moderators to reject papers that are “unrefereable,” contain excessive overlap with prior work, or lie outside the accepted scope of a subject class. In practice, these criteria are ambiguous enough to allow considerable room for subjective judgment.

Historical incidents illustrate the potential for overreach. In 2016, the number theorist Jesús Guillera publicly documented a case in which several of his papers on Ramanujan-type series were repeatedly rejected by mathematics moderators, despite having been published in peer-reviewed journals, because they were deemed to fall outside the moderator’s personal definition of admissible content [17, 18]. The episode became a well-known cautionary tale about “keyword-based” moderation, in which papers are rejected not for lack of rigor but for us-

ing vocabulary or approaches unfamiliar to a moderator. Other researchers have reported similar experiences, particularly in fields where established paradigms are contested [19, 20]. These accounts, while anecdotal, point to a structural vulnerability: when moderation criteria are not operationalized in a publicly auditable form, they can become vectors for the personal idiosyncrasies of individuals who hold gatekeeping positions.

The challenge of maintaining editorial standards is about to be compounded by the rapid maturation of generative artificial intelligence tools capable of producing manuscripts that are stylistically and structurally indistinguishable from those written by expert researchers. When the cost of producing a plausible preprint approaches zero, moderation can no longer rely on surface-level indicators of scientific rigor. The system must either accept a potentially overwhelming volume of submissions or adopt criteria that reach beyond the text itself. That shift opens the door to gatekeeping based not only on author identity and institutional affiliation, but also on strategic interests—commercial, geopolitical, or ideological—that may bear no relation to the scientific merit of the work and need not be declared to be effective. The very opacity of the new governance structure makes such a drift both more feasible and more difficult to detect.

In disciplines where priority of discovery carries high stakes, the absence of binding moderation timelines introduces a structural vulnerability that requires no bad faith to operate. A submission delayed without explanation for weeks or months leaves its author unable to circulate the result widely with the imprimatur of the repository, while competing groups can prepare responses, submit rival preprints, or shape the narrative before the original work becomes publicly visible. The author, frozen in procedural limbo, loses the initiative without ever receiving a formal rejection. Whether such delays ever occur by design is impossible to ascertain without transparent records; that the current system renders the question unanswerable is, in itself, the problem.

If the new governance structure concentrates the appointment of moderators in the hands of a CEO or a small board committee, there is a conceivable risk that moderation practices will drift toward conformity with the interests of the institutions most heavily represented on the board. A board dominated, for example, by researchers from elite U.S. institutions in a specific subfield might, through entirely normal professional networks, tend to select moderators whose sensibilities align with their own, thereby disadvantaging emerging or heterodox lines of inquiry without any conscious intention to do so. Because moderation decisions are currently confidential and not subject to external audit, such a drift could persist undetected for years.

We ask whether the new entity will commit to: (i) a publicly documented moderation code, expressed in operational terms that constrain the use of personal judgment; (ii) an independent ombuds office to handle au-

thor appeals; and (iii) periodic, anonymized release of rejection data by subject class and stated reason, so that patterns suggestive of bias can be detected statistically.

C. Financial Sustainability and the Specter of Submission Fees

The endowment provided by the Simons Foundation, while generous, will not sustain arXiv indefinitely at current growth rates. The organization’s own financial projections, shared in community town hall meetings, indicate that a diversified revenue model will become necessary within five to ten years [7]. Among the options that have been discussed informally are submission fees, article processing charges for certain categories, and increased membership fees for supporting institutions.

The introduction of any mandatory author-facing fee, even if modest, would immediately create a two-tier system. Well-funded research groups in North America and Western Europe would barely notice the cost, while researchers in low-resource settings would face a new and potentially prohibitive barrier to participation. The Global South already contends with multiple structural disadvantages in scientific communication; an author-pays model for preprints would add yet another layer of inequity and would signal a significant departure from the founding principle that arXiv serves as a public utility, not a marketplace.

The board has not yet issued a binding public statement that submission will remain free in perpetuity for all qualified researchers. Without such a commitment, the community is left to speculate about future scenarios. We ask: (i) Will the board adopt a charter that explicitly prohibits mandatory submission fees for primary authors from low- and middle-income countries? (ii) What mechanisms will ensure that institutional membership contributions do not evolve into a governance tier in which paying members gain privileged access to editorial or visibility functions?

D. Systemic Bias: The Unseen Architecture of Exclusion

Systemic bias in scholarly communication is extensively documented. Peer review, journal editorial boards, and citation networks all exhibit persistent disparities correlated with gender, geography, and institutional prestige [22–24]. Preprint repositories were long assumed to be more democratic, because they bypass traditional editorial gatekeeping. However, recent work has demonstrated that even in ostensibly open platforms, participation patterns remain skewed: submissions from low-GDP countries are disproportionately few, and the citation advantage conferred by a preprint is modulated by the author’s institutional affiliation [25, 26].

The governance transition presents a risk that these

disparities will become structurally embedded in arXiv’s decision-making processes. If the board and the moderation corps are drawn predominantly from a narrow set of elite institutions, the repository may gradually optimize for the needs, communication styles, and intellectual preferences of that constituency, reproducing within a preprint repository the same exclusionary dynamics that the open-access movement was created to dismantle. This is not a problem that market forces will correct, because users cannot easily switch to an alternative without fragmenting the literature on which their careers depend. arXiv enjoys a near-monopoly in several disciplines, and network effects lock in that position.

Generative artificial intelligence, for all its democratizing potential, risks accentuating these disparities if governance structures are not designed to anticipate them. A researcher with a deep but heterodox insight who lacks the formal training to package it for elite consumption can now enlist a GAI to co-write a paper that passes surface scrutiny. This is a positive development only if the moderation system judges the work on its merits. If, instead, the system responds to the increased submission volume by using institutional affiliation or professional pedigree as a shortcut for credibility, the very tool that could broaden participation will instead reinforce the epistemic monopoly of incumbent institutions.

We ask: (i) Will the new entity collect and publish aggregated demographic data on its moderator pool and advisory bodies, while respecting individual privacy? (ii) Will it institute a regular equity audit, modeled on those increasingly adopted by scientific societies? (iii) Will there be deliberate efforts to recruit moderators and board members from regions and institution types that are currently underrepresented?

E. Scientific Pluralism and the Boundary Problem

One of the perennial justifications for arXiv moderation is the need to filter out content described by the guidelines as “of a non-scientific nature.” The practical challenge is that the boundary between speculative science and non-science is historically contingent. Ideas now regarded as foundational—from Wegener’s continental drift to the existence of quasicrystals—were initially dismissed as fringe by mainstream gatekeepers [27, 28]. A moderation system optimized to suppress noise may, over time, also suppress signal with an unpredictable delay.

The new governance model, with its professionalized leadership and its sensitivity to institutional reputation, will face structural pressure to avoid controversy. That pressure can translate into overly conservative moderation, in which any paper that challenges a well-entrenched paradigm is more likely to be delayed or rejected than a paper that incrementally extends it. This concern has been voiced repeatedly in fields such as gravitation and cosmology, where researchers pursuing

alternative models occasionally report difficulty in passing the moderation filter [19–21].

When confidence in a central repository’s governance erodes, the scholarly literature does not homogenize; it fragments into alternative servers with less rigorous quality standards, a phenomenon already observed in the last decade that ultimately undermines the very curatorial function arXiv wishes to preserve. This fragmentation is not a theoretical scenario but an empirically observable consequence of opaque moderation, and it serves as a structural warning: the more arXiv is perceived as an exclusive gatekeeper rather than a transparent steward, the more the community will seek outlets that, while preserving openness, lack the quality safeguards that benefit the entire scientific ecosystem.

We do not advocate the elimination of quality control; some filtering is necessary to keep the repository navigable. But we argue that the criteria should be constrained to procedural, testable standards—such as whether the paper presents a clearly stated question, a recognizable methodology, and a logical chain of reasoning—rather than to substantive judgments about the plausibility of the conclusion. We ask whether the new board will endorse such a procedural standard in a public moderation charter.

F. Historical Lessons from Academic Publishing

It is instructive to recall that the modern system of commercial journal publishing began in a structurally similar fashion. Many society journals were initially created by academic communities to serve their own needs, with governance that was lean and mission-driven. Over time, as costs grew and professional management was introduced, some societies outsourced publication to commercial partners, who eventually acquired ownership and converted the journals into high-margin products [5, 8, 9]. The resulting market concentration, in which a handful of publishers control a majority of the scientific record, is now widely recognized as detrimental to both science and the public interest.

arXiv is not a journal, but its role as a primary dissemination platform gives it a comparable centrality in the intellectual life of its user communities. If the new independent entity were ever to face a severe financial crisis, the temptation to monetize the platform—through tiered access, author fees, data licensing, or other commercial arrangements—could prove difficult to resist, particularly if the board includes members with fiduciary ties to the scholarly publishing industry. The most effective defense against such a trajectory is to embed the public-interest mission into the legal charter of the organization in a form that cannot be altered without broad community consent. To our knowledge, such a charter provision has not been announced.

It is no coincidence that the arXiv governance transition is occurring precisely as generative artificial intelli-

gence begins to alter the material conditions of scientific authorship. Historical episodes of infrastructure privatization frequently coincided with moments of technological disruption, when incumbent institutions argued that only professionalized, centralized management could handle the new complexity. The risk in such moments is that structures designed to manage a crisis become permanent features that long outlast the conditions that produced them. The governance choices made in 2026 will determine whether arXiv becomes a platform that harnesses artificial intelligence to broaden the scientific conversation or one that uses it as a pretext to narrow it.

V. SAFEGUARDING THE PUBLIC INTEREST: A COMMUNITY-CENTERED GOVERNANCE FRAMEWORK

The risks identified in the preceding section are structural, and their mitigation requires structural solutions. We propose below a set of five concrete mechanisms that, if adopted by the board of arXiv Inc., would substantially increase transparency, accountability, and community participation without impeding the operational efficiency of the repository.

1. **Community Advisory Council.** A standing council with members elected by the global user community from disciplinary and geographic constituencies should be formally integrated into the governance structure. The council would have the right to review moderation policies, receive anonymized data on submission and acceptance rates, and issue non-binding but public recommendations to the board. Its existence should be codified in the organizational bylaws so that it cannot be dissolved at the discretion of the board alone.
2. **Transparency Portal.** The organization should maintain a publicly accessible web portal containing, at a minimum: the board’s bylaws and meeting minutes; annual financial statements with disclosure of donors above a specified threshold; monthly submission and rejection statistics per subject class; and an aggregated demographic profile of the moderator corps. This portal would serve as the primary instrument of public accountability.
3. **Independent Moderation Ombuds.** An ombudsperson, appointed for a fixed term and insulated from removal without cause, should be empowered to handle appeals from authors who believe their submission was rejected in violation of the published moderation code. The ombuds’ findings, suitably anonymized, should be made public to establish a body of precedent that constrains moderator discretion over time.
4. **Free Access Charter.** The organizational character should include a clause explicitly prohibiting

mandatory submission or access fees for individual researchers. Should the board ever seek to introduce such fees, a supermajority vote of the Community Advisory Council and a public consultation period of no less than six months should be required.

5. **Periodic External Audit.** Every five years, an independent panel appointed jointly by the board and the Community Advisory Council should conduct a comprehensive audit of moderation practices, financial management, equity outcomes, and user satisfaction, publishing its full report in the transparency portal.

These proposals are not unprecedented. They draw on governance models already operational in institutions such as the Wikimedia Foundation, the OpenAIRE infrastructure, the Public Library of Science, and the Open Science Framework [31–33]. What makes them necessary in this context is the unique quasi-monopoly position of arXiv in several disciplines and the near-total absence of formal accountability mechanisms in its publicly described new design.

VI. CONCLUSIONS

The transition of arXiv to an independent nonprofit organization occurs at a moment when the role of preprint servers in the scientific ecosystem has never been more prominent. The repository processes nearly twenty thousand new submissions each month, serves a global and increasingly multidisciplinary readership, and shapes the visibility and career impact of research in fields that span the quantitative sciences. Its governance is therefore a matter of legitimate public concern, not merely an internal organizational matter for Cornell University and its philanthropic partners.

We have argued, on the basis of publicly available evidence and documented historical episodes, that the transition plan as currently described lacks the transparency and community-accountability provisions necessary to safeguard arXiv’s public-interest mission over the long term. The board’s composition and operational procedures are not constitutionally guaranteed to be transparent, regardless of who currently serves; the future of the moderation system is undefined; the financial model beyond the initial endowment is opaque; and there are no built-in mechanisms to detect, correct, or even measure systemic bias. History offers ample warning that infrastructure of this kind, once separated from its original academic custodianship and subjected to professionalized management without robust community oversight, can drift toward private capture, even when all parties involved initially act with the best intentions.

The moment before the new governance structure becomes operationally entrenched is precisely the moment when it is easiest to incorporate binding safeguards. We

therefore urge the board of arXiv Inc. to adopt the transparency and accountability measures outlined in this paper, and we urge scientific societies, funding agencies, and library consortia to condition their financial and institutional support on verifiable commitments to these principles. The arXiv was built by the research community, for the research community. Its future governance should reflect that origin.

As a modest empirical contribution to the debate this article initiates, we have submitted the manuscript to arXiv itself. A single submission cannot, of course, confirm or refute the existence of systemic bias in the moderation process. What it can do is provide a publicly documented data point on the consistency and transparency of the procedure: whether the moderation code is applied, whether a reasoned justification is provided for the editorial decision, and whether the timeline is comparable to that experienced by other submissions. These are the procedural metrics that matter, and we invite the community to track them alongside us. The guardians of the preprint republic now face a choice that Plato’s original guardians never had to confront: whether the legitimacy of their rule should rest on authority alone, or on the transparent consent of the governed.

LIMITATIONS

This analysis is subject to several limitations. First, it relies primarily on publicly available evidence—submission statistics, official announcements, news reports, and documented moderation controversies—rather than on internal arXiv data. A comprehensive audit of moderation patterns, including acceptance rates stratified by author demographics and institutional affiliations, would require access to data that only arXiv Inc. can provide. Our conclusions should therefore be read as a structural risk assessment, not as an empirical demonstration of systemic failure. Second, the documented moderation controversies we cite are limited in number and may not represent the typical experience of arXiv users; they serve as existence proofs of vulnerability rather than as statistical evidence of widespread bias. Third, the governance transition is still in progress, and some of the transparency deficits we identify may be resolved before the new entity becomes operational. We hope that this article contributes to that resolution. Fourth, our analysis of the implications of generative artificial intelligence for moderation is necessarily speculative; the technology is evolving rapidly, and its effects on submission volume and quality are not yet fully measurable. Future research should combine internal arXiv data, user surveys, and longitudinal analysis of moderation outcomes to test the hypotheses advanced here.

-
- [1] P. Ginsparg, “First steps towards electronic research communication,” *Computers in Physics* **8**, 390–396 (1994).
 - [2] P. Ginsparg, “It was twenty years ago today...”, arXiv:1108.2700 (2011).
 - [3] arXiv.org, “Monthly submission statistics,” https://arxiv.org/stats/monthly_submissions (accessed July 2026).
 - [4] R. J. Adler, P. Etingof, and D. Vogan, “The arXiv at 30,” *Notices of the American Mathematical Society* **68**, 1947–1952 (2021).
 - [5] V. Larivière, S. Haustein, and P. Mongeon, “The oligopoly of academic publishers in the digital era,” *PLoS ONE* **10**, e0127502 (2015).
 - [6] J. Willinsky, *The Access Principle: The Case for Open Access to Research and Scholarship* (MIT Press, Cambridge, 2006).
 - [7] Cornell University Library, “arXiv’s future: becoming an independent nonprofit,” press release, May 2025.
 - [8] M. J. McCabe, “Journal pricing and mergers: a portfolio approach,” *American Economic Review* **92**, 259–263 (2002).
 - [9] P. Suber, *Open Access* (MIT Press, Cambridge, 2012).
 - [10] V. Larivière, C. Sugimoto, and B. Macaluso, “The arXiv as a vehicle for accelerated diffusion,” *Journal of the Association for Information Science and Technology* **65**, 1165–1177 (2014).
 - [11] C. Brown, “The role of electronic preprints in chemical communication: analysis of citation and usage patterns,” *Journal of Chemical Information and Modeling* **47**, 1107–1112 (2007).
 - [12] C. Soderberg, S. E. Hobson, and M. P. Peterson, “Preprints and the speed of science,” *Scientometrics* **124**, 2375–2390 (2020).
 - [13] N. Fraser *et al.*, “The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape,” *PLoS Biology* **19**, e3000959 (2021).
 - [14] C. Watson, “Preprints and pandemic science: lessons learned,” *Learned Publishing* **36**, 124–130 (2023).
 - [15] T. Schillemans, *Accountability in the Shadow of Hierarchy: The Horizontal Accountability of Agencies* (Springer, 2010).
 - [16] A. J. Meijer, “Transparency,” in *The Oxford Handbook of Public Accountability*, edited by M. Bovens, R. E. Goodin, and T. Schillemans (Oxford University Press, 2016).
 - [17] J. Guillerá, “On the arXiv moderation of papers on Ramanujan’s mathematics,” preprint (2016). Available at <https://www.researchgate.net/publication/311557409>.
 - [18] J. C. Baez, “The arXiv moderation crisis,” Azimuth blog (2017).
 - [19] S. Hossenfelder, *Lost in Math: How Beauty Leads Physics Astray* (Basic Books, 2018).
 - [20] L. Smolin, *The Trouble with Physics* (Houghton Mifflin, 2006).

- [21] L. Smolin, *The Singular Universe and the Reality of Time* (Penguin, 2015).
- [22] V. Larivière *et al.*, “Bibliometrics: global gender disparities in science,” *Nature* **504**, 211–213 (2013).
- [23] C. J. Lee *et al.*, “Bias in peer review,” *Journal of the American Society for Information Science and Technology* **64**, 2–17 (2013).
- [24] M. M. King *et al.*, “Gender and geographical disparity in editorial boards of scientific journals,” *Nature Human Behaviour* **6**, 1012–1021 (2022).
- [25] G. Abramo, C. A. D’Angelo, and F. Di Costa, “The effects of gender, age and academic rank on research diversification,” *Scientometrics* **127**, 2575–2596 (2022).
- [26] Y. Xie, K. Wang, and J. You, “The stratified effect of preprints on citation impact,” *Quantitative Science Studies* **4**, 315–331 (2023).
- [27] I. Langmuir, “Pathological science,” *Physics Today* **42**, 36–48 (1989). [Reprint of 1953 lecture.]
- [28] R. W. Cahn, “How quasicrystals were discovered,” *Nature Materials* **10**, 724–725 (2011).
- [29] S. Pinfield *et al.*, “Open-access repositories worldwide, 2005–2012: past growth, current characteristics, and future possibilities,” *Journal of the Association for Information Science and Technology* **65**, 2404–2421 (2014).
- [30] Simons Foundation, “Grant to arXiv: ensuring the future of open access preprints,” press release, 2025.
- [31] Public Library of Science, “Bylaws and governance documents,” <https://plos.org/about/governance/> (accessed July 2026).
- [32] Wikimedia Foundation, “Board of Trustees governance handbook,” 2020.
- [33] E. D. Foster and B. Johnson, “Open Science Framework (OSF),” *Journal of the Medical Library Association* **105**, 203–206 (2017).
- [34] The Simons Foundation is a private philanthropic organization established in 1994 by Jim Simons, a mathematician and former hedge fund manager, and his wife Marilyn Simons. It is among the largest private funders of basic science in the United States.