Editorial

Artificial intelligence in scientific publishing: functional alegality and epistemic inequality

Angel Gordo
TRANSOC, Universidad Complutense de Madrid, Spain.
ajgordol@ucm.es

Pompeu Casanovas Romeu
Instituto de Investigación en Inteligencia Artificial (IIIA-CSIC), Spain. La Trobe University, Australia.
pompeu.casanovas@iiia.csic.es

ABSTRACT

The silent incorporation of artificial intelligence into editorial processes is reshaping the ecosystem of scientific publishing. In this editorial, we argue that this incorporation operates within a condition of functional alegality, where editorial actors deploy AI without explicit regulation, traceability, or accountability. This process widens existing asymmetries between large publishing consortia and diamond open access journals. We examine how algorithmic biases reproduce epistemic inequalities and how AI intensifies, rather than resolves, the structural crisis of peer review. In response, we defend a governance framework grounded in compliance through design, traceability, and collective deliberation. AI can either rebalance scientific publishing or deepen concentration. The outcome depends on who governs its use.

Keywords: artificial intelligence; scientific publishing; compliance through design; diamond open access; epistemic justice; peer review; open science.

***Corresponding author: Ángel Gordo,** ajgordol@ucm.es

Suggested citation: Gordo, Á. y Casanovas Romeu, P. (2026). Artificial intelligence in scientific publishing: functional alegality and epistemic inequality. *Revista Española de Sociología*, *35*(3), a305. https://doi.org/10.22325/fes/res.2026.305

The emergence of generative AI in editorial processes has reconfigured the ecosystem of scientific publishing. Although journals, platforms, and manuscript management systems rarely acknowledge it explicitly, language model–assisted review has already become routine in many editorial settings. The promise of reducing reviewer overload, accelerating decision-making, and standardizing reports has helped legitimize its adoption. Its immediate antecedents lie in the machine learning systems that, over the last decade, have supported the evaluation of submissions and posters at large-scale international conferences such as the International Conference on Machine Learning (ICML), the Conference on Neural Information Processing Systems (NeurIPS), and the International Conference on Learning Representations (ICLR), where submission volumes made exclusively human review unviable (Kuznetsov et al., 2024).

AI does not merely accelerate scientific review; it redefines decision-making processes and shifts the criteria for what counts as valid knowledge (Naddaf, 2025). The central issue is not the speed of evaluation but its meaning and, with it, the implicit social contract underpinning trust in science. This transformation raises a critical question: does the silent normalization of AI in editorial decisions (and, by extension, in public calls for proposals and administrative procedures) compromise the social consensus on which scientific credibility rests? It also raises a second question: do current peer review systems, which scholars have questioned for decades (Tennant et al., 2017), possess enough legitimacy to withstand this transformation, or can they adapt to the new dynamics of knowledge production and evaluation generated by generative AI (Thelwall, 2025)?

These questions are not merely theoretical. Their consequences appear in the growing gaps that the silent incorporation of AI creates between different publication models. Large international consortia, including Springer Nature, Elsevier, and Wiley, operate through closed, fee-based systems that impose high costs on both authors and readers. By contrast, independent or institutional academic journals generally operate through open access models that charge no fees to authors or readers, the so-called diamond open access journals. The incorporation of AI into editorial processes intensifies this preexisting structural asymmetry, particularly when publishers deploy these tools opaquely, without explicit regulation, accountability, or the informed consent of the parties involved.

Publishing consortia have quickly recognized that AI is not merely a technical resource but a strategic infrastructure for reinforcing their control over knowledge flows. Automated systems have long handled initial screening tasks, including so-called desk rejections, formal compliance checks, and plagiarism detection. With the arrival of general-purpose generative models, this automation now extends to decisions that, until recently, required expert human judgment. Tools such as MetaWriter and ReviewFlow can draft peer review reports, generate editorial summaries from disparate evaluations, and synthesize observations with a versatility that, according to some studies, approaches human editorial work (Kankanhalli, 2024).

Although publishers proclaim their commitment to ethics and transparency, few explain how they use AI in their internal processes. The COPE report acknowledges that most editors rely on AI for screening, fraud detection, or reviewer suggestions without explicitly informing authors or the public (Zhou & Soulière, 2025). This gray zone constitutes what we define as functional alegality: a space of practice that does not violate explicit norms because those norms do not yet exist, yet still operates outside mechanisms of scrutiny, traceability, and accountability. This is a self-serving form of anomie, not only accepted or maintained, but often directly sought after or even fostered. The normative discourse appeals to transparency and intellectual honesty, while the system relies on proprietary architectures that reproduce the very inequalities those appeals claim to correct.

The rhetoric of human oversight at all stages of the process, commonly invoked through the formula of human-in-the-loop, rarely comes with concrete procedural mechanisms that allow authors to demand accountability in the event of a dispute. Empirical evidence confirms that transparency improves the quality of scientific evaluation itself. According to the recent study by Álvarez-García et al. (2026), openly published peer review reports contain substantially more informational content than confidential reports and focus more strongly on constructive suggestions. Organizations such as the Association for Computing Machinery (ACM, 2023) and the CSIC (2026) agree that algorithmic systems cannot guarantee reliability without clear mechanisms for challenge and redress (Makarem et al., 2023; Baeza-Yates, 2024). However, authors rarely exercise that right beyond an immediate rebuttal of a rejection—a practical resignation that normalizes opacity and erodes the system's legitimacy from within.

Specialized literature has documented the class, ethnicity, language, and institutional biases that structure algorithmic systems. In the editorial domain, these biases not only persist but also intensify. Farber (2024, p. 6) shows that journals with the resources to implement AI in review significantly reduce their decision times while increasing the acceptance rate of articles from high-prestige institutions. Algorithmic bias and structural biases reinforce each other. Models learn from dominant corpora, reproduce their hierarchies, and penalize approaches that deviate from those patterns (Mühlhoff, 2025). Work from peripheral regions or critical traditions therefore faces a double invisibility: the invisibility produced by metrics and the invisibility produced by algorithms. As Thelwall (2019) notes, the automation of scientific evaluation tends to reinforce asymmetries of academic capital and erode epistemic diversity under the guise of technical neutrality.

Artificial intelligence, even in its generative form, is not an autonomous entity but the sedimentation of human decisions encoded in systems whose traceability often remains opaque or directly inaccessible. Models learn from the corpora we provide, reproduce our biases, and amplify our hierarchies. As Ebadi et al. (2025) show, even reviewer recommendation systems tend to favor authors from Anglophone institutions and penalize linguistic diversity. Certain high-frequency patterns in pre-training data can also distort model learning, introducing biases that developers cannot always detect or correct (Zevallos, Farrús, & Bel, 2023). Training bias thus becomes legitimation bias. For this reason, editorial data (from reviewer databases to article repositories) must be open, plural, and representative. Data openness is not only a technical imperative but also a democratic principle, including in post-publication evaluation, where articles can remain subject to ongoing review and collective assessment.

The expansion of AI has also occurred amid a structural crisis in peer review, marked by reviewer fatigue, precarious editorial labor, and the massive increase in submissions. Hosseini and Horbach (2023) note that the annual volume of manuscripts has doubled over the last decade, while the availability of qualified reviewers has grown by less than 20 percent. In Spanish sociology, the *Revista Española de Sociología* (Spanish Journal of Sociology) tripled the number of manuscripts received between 2017 and 2023 (Cárdenas, 2023), a trend that reflects broader pressures in the social science publishing ecosystem. This mismatch has consequences beyond editorial logistics: it erodes the model of scientific community on which peer review once rested. In the early years of this century, before the expansion of the publishing market and the proliferation of journals driven more by investment logics than by knowledge production, scholarly communities themselves supplied reviewers with the necessary expertise and a genuine interest in the state of the art in their fields. Over the last fifteen years, that dynamic has changed radically. Journals now contact reviewers with whom they have no prior relationship, publish on increasingly compressed timelines, and incentivize a competitive culture that pressures young researchers to publish before they have consolidated their training.

Faced with this mismatch, AI presents itself as a pragmatic solution, a technological patch for a systemic problem. Yet what editorial systems gain in efficiency, they often lose in deliberation and epistemic plurality. The homogenization of evaluation tends to displace the diversity of disciplinary criteria and, in doing so, impoverishes the conditions under which knowledge is produced. This is not a speculative hypothesis: scholars have already linked the decline in the scientific quality of publications to the uncritical use of AI (Naddaf, 2025).

In this scenario, large publishing consortia experiment with AI under the rhetoric of innovation and sustainability, but they rely on closed models and privatized data. Independent and public journals, often operating as diamond open access journals, face a different dilemma. They must decide whether to use AI while lacking the material resources to do so on their own terms. Without the capacity to develop specific models and under pressure from indexing criteria, they oscillate between abstention and principled rejection. Yet the gap is not only economic or infrastructural; it is also disciplinary. The experimental sciences, accustomed to procedural standardization and intensive computational tools, have adapted relatively quickly to the new centrality of AI. The social sciences and humanities, which depend more heavily on interpretive judgment and contextual analysis, tend to respond with greater caution or outright rejection. The technological gap thus becomes an epistemic gap. Empirical evidence shows that social science journals maintain the highest standards in the deliberative function of peer review, surpassing both the experimental sciences and the health sciences (García-Costa et al., 2022). Their epistemic specificity is not a limitation but their principal asset in this debate. Journals in these disciplines therefore face a double challenge: they must avoid technological marginalization while refusing to cede their epistemic sovereignty to models and logics alien to their methodological and interpretive specificity.

The problem is not whether journals should adopt or reject AI, but under what conditions they should do so. Rejecting AI on principle means renouncing participation in debates about its design, ethics, and social function. An ethics of non-use can prove as irresponsible as uncritical use, because both extremes perpetuate inequality. Independent journals must therefore do more than ignore AI; they must build governance frameworks that subject it to collective deliberation and public oversight. Large consortia and technology corporations will not provide this response. It must come from scientific communities capable of developing alternatives.

Peer review emerged as a human system of trust, not as an algorithmic mechanism of validation. Its current crisis reflects not only manuscript saturation but also the erosion of that trust. AI can worsen this crisis if editorial systems use it as a substitute for human responsibility. However, AI can also help restore trust if journals integrate it according to ethical and transparent criteria. The dilemma is not technical but governmental: who defines the rules, who audits the algorithms, and who benefits from their use. Drawing on the distinction between regulation, which treats agents (human or not) as autonomous actors capable of complying with or transgressing norms, and regimentation, which reduces autonomy by programming compliance into system design, Casanovas (2025) proposes a hybrid strategy that combines social architecture with technical architecture. The objective is not to impose external norms but to integrate ethics into system design, so that transparency, traceability, and human oversight become constitutive properties of these tools rather than add-ons appended later (Casanovas et al., 2025; Noriega et al., 2023).

This perspective connects with the value alignment problem: how to ensure that AI systems act in accordance with the values that the communities using them seek to preserve. Russell (2023) provides the most influential formulation of this problem. He argues that AI systems should not optimize fixed objectives but should instead operate with epistemic humility regarding human preferences and retain the capacity to defer to human judgment—

an idea that underlies the governance frameworks proposed here (Osman & Steels, 2024; Casanovas & Noriega, 2026).

This governance horizon corresponds to regulatory frameworks such as the European AI Act (European Parliament and Council of the European Union, 2024) and finds a concrete, though still developing, expression in the multilevel compliance model proposed in Gordo et al. (2025). Its central principle holds that transparency, traceability, and human oversight must constitute intrinsic properties of AI tools, not supplementary features added later. More precisely, the authors propose the concept of compliance through design: this approach does not program compliance with fixed rules but instead designs systems flexible enough to incorporate different interpretations of the values that underpin them (Casanovas et al., 2025). This model includes six levels: shared academic criteria, modeling and training, scientific integrity, ethical safeguards, institutional governance, and sustainable deployment. These levels do not form a hierarchy; they operate as a system of feedback loops. Each level incorporates mechanisms of traceability and human oversight that encompass not only data but also editorial decisions and algorithmic interventions. In this model, editors document every automated suggestion, identify and correct every error, and submit every decision to collective discussion. Applied to peer review, this approach means that editorial sustainability cannot be reduced to faster timelines or more efficient management processes. Instead, journals must understand sustainability as a form of epistemic justice. The model proposes assisted deliberation, not substitution.

A model of this kind could transform the ecosystem of diamond open access journals. It would reduce dependence on closed commercial platforms and support the construction of open and interoperable infrastructures. Commercial publishers tend to convert each technological advance into a paid service. Against this tendency, alliances among universities, open-source software platforms, and editorial communities could articulate an alternative model of shared editorial intelligence in which AI functions as shared infrastructure rather than as a proprietary product, while collective deliberation prevails over private optimization. From this perspective, AI ceases to appear only as a threat and becomes an instrument for rebalancing asymmetrical capacities across publishing models. Achieving that balance, however, requires political will, institutional coordination, and a culture of cooperation that remains underdeveloped in much of academia.

Building this horizon requires open, auditable, and collectively governed infrastructures that guarantee scientific communities control over the processes of knowledge evaluation and knowledge validation. Unlike commercial platforms that concentrate power, the multilevel compliance approach promotes distributed practices and technologies in which AI supports deliberation rather than replaces it. The aim is not to humanize AI, as though it were an external entity that needed values added to it, but to socialize its design by orienting it toward the principles of cooperation, transparency, and accountability that underpin open science.

Scientific communities can no longer delegate the question of governance. They must recover decision-making capacity over the tools they use, which requires technical training, institutional frameworks for cooperation, and political will. European guidelines recommend that researchers retain ultimate responsibility for AI-assisted generated content, but individual responsibility remains insufficient without collective structures of oversight and participation. An ethics of human agency must therefore extend toward an ethics of the community. It is not enough for each researcher to remain accountable for individual uses of AI if the system as a whole does not undergo collective deliberation. At this point, diamond open access journals play a fundamental role. Because of their public character and economic independence, they can become laboratories of distributed governance: spaces where scientific communities build and consolidate alternative forms of open editorial intelligence.

The democratization of AI in scientific publishing will not come from declarations of principle. It will emerge from the daily construction of transparent and verifiable structures. Oversight must be real, errors must be correctable, and decisions must be contestable. In this new context, scientific integrity does not require prohibiting AI use. It requires ensuring that AI intervention remains visible, contextualized, and subject to collective deliberation. Scientific knowledge has always depended on technical mediations. What distinguishes AI is its capacity to classify, evaluate, and ultimately decide. Sociological critique must therefore shift from the tool to the system, and from the algorithm to the power regime that sustains it.

The AI gap in scientific publishing is not simply a technological imbalance. It is a matter of values and collective choice about the model of science we want to sustain. Closing that gap does not require renouncing technology; it requires restoring technology to collective control. Journals, editors, reviewers, authors, and readers all participate in the same process of legitimation, which can only rest on trust and transparency.

Redefining this process of legitimation requires returning to a fundamental question. As Irfanullah (2023) reminds us, peer review emerged as a human system of trust, not as an algorithmic mechanism of validation. Trust cannot be programmed; communities must build it through transparency and reciprocity. Faced with the consolidation of AI in editorial processes, the question is no longer whether scientific publishing should accept AI, but how it can govern AI without losing the public and deliberative dimensions of knowledge. The answer requires recognizing that automation does not replace responsibility and that efficiency cannot serve as the sole criterion of value. If scientific communities subject artificial intelligence to these principles, AI can become an ally in recovering the public meaning of science. If they abandon it to the market, it will become another instrument of concentration and exclusion.

REFERENCES

Álvarez-García, E., García-Costa, D., Squazzoni, F., Malički, M., Mehmani, B., & Grimaldo, F. (2026). Published peer review reports have higher informative content than unpublished reports. *Journal of Informetrics*, *20*(1), 101760. https://doi.org/10.1016/j.joi.2025.101760

Association for Computing Machinery [ACM] (2023, June 27th). *Principles for the development, deployment, and use of generative AI technologies*. ACM Technology Policy Office. https://www.acm.org/binaries/content/assets/public-policy/ustpc-approved-generative-ai-principles

Baeza-Yates, R. (2024). Recomendaciones para una IA responsable. *Cuadernos de Beauchef*, *8*(1), 99-111. https://revistasdex.uchile.cl/index.php/cdb/article/view/15230/15264

Cárdenas, J. (2023). Inteligencia artificial, investigación y revisión por pares: Escenarios futuros y estrategias de acción. *Revista Española de Sociología*, *32*(4), a184. https://doi.org/10.22325/fes/res.2023.184

Casanovas, P. (2025). La regimentación tecnológica: Inteligencia artificial, fascismo, agresión y sociedad democrática. *Teknokultura: Revista de Cultura Digital y Movimientos Sociales*, *22*(2), 137-149. https://doi.org/10.5209/tekn.98266

Casanovas, P., & Noriega, P. (2026). Governance of Artificial Agency and AI Value Chains: A Few Remarks on Autonomy from a Legal and Ethical Approach. In D. López Castro, M. Cebral, & P. Jiménez-Schlegl, *Regulating Autonomy: Ethics, Values and Governance in Artificial Intelligence*. Springer Nature. https://doi.org/10.1007/978-3-032-13063-1_16

Casanovas, P., Hashmi, M., De Koker, L., & Lam, H. P. (2025). Compliance, Regtech, and smart legal ecosystems: a methodology for legal governance validation. In W. Barfield, U. Pagallo (eds.) *Research handbook on the law of artificial intelligence. Current and Future Directions: Second Edition* (pp. 73-104). Edward Elgar Publishing. https://doi.org/10.4337/9781035316496.00011

Consejo Superior de Investigaciones Científicas [CSIC] (2026). *Decálogo IA*. https://www.csic.es/es/decalogo-ia

Ebadi, S., Nejadghanbar, H., Salman, A. R., & Khosravi, H. (2025). Exploring the impact of generative AI on peer review: Insights from journal reviewers. *Journal of Academic Ethics*, *23*(3), 1383-1397. https://doi.org/10.1007/s10805-025-09604-4

Farber, S. (2024). Enhancing Peer Review Efficiency: A Mixed-Methods Analysis of Artificial Intelligence-Assisted Reviewer Selection Across Academic Disciplines. *Learned Publishing*, 37, e1638. https://doi.org/10.1002/leap.1638

García-Costa, D., Squazzoni, F., Mehmani, B., & Grimaldo, F. (2022). Measuring the developmental function of peer review: A multi-dimensional, cross-disciplinary analysis of peer review reports from 740 academic journals. *PeerJ*, *10*, e13539. https://doi.org/10.7717/peerj.13539

Gordo, A., Casanovas, P., Méndez, E., Rodríguez, A., Sierra, C., Cruz Mencía, F., Lemus del Cueto, L., Poblet, M., López Tamayo, E., González Guardia, E., Jené, J., Said Hung, E., Tabarés, R., Garijo, D., Rodríguez, V., Gómez, A., Montiel, E., Osman, N., Sokolowicz Papiol, N., & Cester Mazarico, L. (2025, July 4th). *Aligning Generative AI with Scientific Responsibility: A Multi-Level Compliance Model for Peer Review.* [Unpublished white paper]

Hosseini, M., & Horbach, S. P. (2023). Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. *Research Integrity and Peer Review*, *8*(1), 4. https://doi.org/10.1186/s41073-023-00133-5

Irfanullah, H. (2023, September 29th). Ending Human-Dependent Peer Review. *The Scholarly Kitchen*. https://scholarlykitchen.sspnet.org/2023/09/29/ending-human-dependent-peer-review/

Kankanhalli, A. (2024). Peer Review in the Age of Generative AI. *Journal of the Association for Information Systems*, *25*(1), 76-84. https://doi.org/10.17705/1jais.00865

Kuznetsov, I., Afzal, O. M., Dercksen, K., Dycke, N., Goldberg, A., Hope, T., Hovy, D., Kummerfeld, J. K., Lauscher, A., Leyton-Brown, K., Lu, S., Mausam, Mieskes, M., Névéol, A., Pruthi, D., Qu, L., Schwartz, R., Smith, N. A., Solorio, T., Wang, J., Zhu, X., Rogers, A., Shah, N. B., & Gurevych, I. (2024). *What can natural language processing do for peer review?* arXiv. https://doi.org/10.48550/arXiv.2405.06563

Makarem, A., Mroué, R., Makarem, H., Diab, L., Hassan, B., Khabsa, J., & Akl, E. A. (2023). Conflict of interest in the peer review process: A survey of peer review reports. *PLoS One*, *18*(6), e0286908. https://doi.org/10.1371/journal.pone.0286908

Mühlhoff, R. (2025). *The ethics of AI: Power, critique, responsibility*. Bristol University Press.

Naddaf, M. (2025, May 29th). AI linked to explosion in low-quality biomedical research papers. *Nature, 641*, 1080-1081. https://doi.org/10.1038/d41586-025-01592-0

Noriega, P., Verhagen, H., Padget, J., & d'Inverno, M. (2023). Coordination, organizations, institutions, norms, and ethics for governance of multi-agent systems. In N. Fornara, J. Cheriyan, & A. Mertzani (Eds.), *Lecture Notes in Computer Science* (Vol 14002, pp. 77-94). Springer. https://doi.org/10.1007/978-3-031-49133-7_5

Osman, N., & Steels, L. (Eds.) (2024). *Value Engineering in Artificial Intelligence*. Second International Workshop, VALE 2024 at IJCAI, Santiago de Compostela, Spain, 19-24 octubre, Revised Selected Papers, LNCS 15356, Springer Nature.

European Parliament and Council of the European Union (2024). *Reglamento (UE) 2024/1689: Ley de Inteligencia Artificial. Diario Oficial de la Unión Europea.* https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=OJ:L_202401689

Russell, S. J. (2023). *Human compatible: Artificial intelligence and the problem of control*. Penguin.

Tennant, J. P., Dugan, J. M., Graziotin, D., Jacques, D. C., Waldner, F., Mietchen, D., Elkhatib, Y., Collister, L. B., Pikas, C. K., Crick, T., Masuzzo, P., Caravaggi, A., Berg, D. R., Niemeyer, K. E., Ross-Hellauer, T., Mannheimer, S., Rigling, L., Katz, D. S., Greshake Tzovaras, B., … & Colomb, J. (2017). *A multi-disciplinary perspective on emergent and future innovations in peer review* (Version 3). *F1000Research, 6*, 1151. https://doi.org/10.12688/f1000research.12037.3

Thelwall, M. (2019). *Artificial intelligence, automation and peer review*. Jisc. https://repository.jisc.ac.uk/7614/1/AI_and_peer_review_briefing_paper.pdf

Thelwall, M. (2025). Research quality evaluation by AI in the era of large language models: Advantages, disadvantages, and systemic effects–An opinion paper. *Scientometrics*, *130*(10), 5309-5321. https://doi.org/10.1007/s11192-025-05361-8

Zevallos, R. J., Farrús, M., & Bel, N. (2023). Frequency balanced datasets lead to better language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 7859-7872). Association for Computational Linguistics. https://aclanthology.org/2023.findings-emnlp.527.pdf

Zhou, H., & Soulière, M. (2025, August 25th). *From Detection to Disclosure—Key Takeaways on AI Ethics from COPE's Forum*. The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2025/08/25/from-detection-to-disclosure-key-takeaways-on-ai-ethics-from-copes-forum/