Editorial / Editorial

La IA en la edición científica: alegalidad funcional y desigualdad epistémica

Artificial intelligence in scientific publishing: functional alegality and epistemic inequality

Ángel Gordo
TRANSOC, Universidad Complutense de Madrid, España.
ajgordol@ucm.es

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

RESUMEN

La incorporación silenciosa de la inteligencia artificial en los procesos editoriales está reconfigurando el ecosistema de la publicación científica. En esta editorial proponemos que esa incorporación opera en una zona de alegalidad funcional, sin regulación explícita ni rendición de cuentas, ampliando las asimetrías entre grandes consorcios editoriales y revistas de ruta diamante. Examinamos los sesgos algorítmicos que reproducen las desigualdades epistémicas existentes y la crisis estructural de la revisión por pares que la IA agrava en lugar de resolver. Frente a estos desafíos, defendemos un marco de gobernanza basado en el cumplimiento incorporado a través del diseño, la trazabilidad y la deliberación colectiva. La IA puede ser un instrumento de reequilibrio o de concentración, y la diferencia depende de quién gobierne su uso.

Palabras clave: inteligencia artificial, publicación científica, cumplimiento normativo mediante el diseño, acceso abierto diamante, justicia epistémica, revisión por pares, ciencia abierta.

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.

***Autor para correspondencia / *Corresponding author:* Ángel Gordo,** ajgordol@ucm.es

Sugerencia de cita / *Suggested citation:* Gordo, Á. y Casanovas Romeu, P. (2026). La IA en la edición científica: alegalidad funcional y desigualdad epistémica. *Revista Española de Sociología, 35*(3), a305. https://doi.org/10.22325/fes/res.2026.305

La irrupción de la IA generativa en los procesos editoriales ha reconfigurado el ecosistema de la publicación científica. Aunque raramente se reconoce de manera explícita, la revisión asistida por modelos de lenguaje forma ya parte de la rutina en un número creciente de revistas, plataformas y sistemas de gestión de manuscritos. La promesa de aliviar la sobrecarga de revisores, agilizar la toma de decisiones y estandarizar los informes ha bastado para legitimar su adopción. Sus antecedentes inmediatos se encuentran en los sistemas de aprendizaje automático que, durante la última década, comenzaron a emplearse para evaluar propuestas de comunicación y pósteres en congresos internacionales de gran escala como la International Conference on Machine Learning (ICML), la Conference on Neural Information Processing Systems (NeurIPS) o la International Conference on Learning Representations (ICLR), donde el volumen de envíos hacía inviable la revisión exclusivamente humana (Kuznetsov et al., 2024).

La IA no se limita a acelerar la revisión científica sino que redefine los procesos de toma de decisión al tiempo que desplaza los criterios de lo que se considera conocimiento válido (Naddaf, 2025). Lo que está en juego no es la velocidad de la evaluación sino su sentido, y con él, el pacto tácito sobre el que se ha sostenido históricamente la confianza en el sistema científico. Esta transformación lleva a plantear si la normalización silenciosa de la IA en las decisiones editoriales (y por extensión, en convocatorias públicas o procedimientos administrativos) no compromete el consenso social sobre el que se edifica la credibilidad científica, o, incluso, si los sistemas vigentes de revisión por pares, cuestionados de forma sostenida durante décadas (Tennant et al., 2017), poseen realmente la legitimidad suficiente para resistir esta transformación, o si están ya a la altura de las nuevas dinámicas de producción y evaluación del conocimiento impulsadas por la IA generativa (Thelwall, 2025).

Estas preguntas no son meramente teóricas. Sus consecuencias se hacen visibles en las brechas que la incorporación silenciosa de la IA está ampliando entre distintos modelos de publicación. Por un lado, las revistas gestionadas por grandes consorcios internacionales (Springer Nature, Elsevier, Wiley) operan con sistemas cerrados y de pago, con costes elevados tanto para quienes publican como para quienes leen. Por otro, las revistas académicas independientes o institucionales funcionan generalmente en acceso abierto y sin cargos económicos para autores/as ni lectores/as, esto es, las llamadas revistas de ruta diamante. Esta asimetría estructural preexistente se intensifica con la incorporación de la IA a los procesos editoriales, y lo hace de manera opaca, sin regulación explícita, sin rendición de cuentas y sin consentimiento informado de ninguna de las partes implicadas.

Los consorcios editoriales no han tardado en comprender que la IA no es solo un recurso técnico, sino una infraestructura estratégica para reforzar su control sobre los flujos de conocimiento. Desde hace tiempo, las tareas de cribado inicial (las llamadas *desk rejections*), basadas en verificaciones de cumplimiento formal y detección de plagio, están a cargo de sistemas automáticos. Con la llegada de los modelos generativos de uso general, esta automatización ha pasado a formar parte de decisiones que hasta hace poco requerían el juicio experto humano. Herramientas como MetaWriter o ReviewFlow permiten redactar borradores de revisión, generar informes editoriales a partir de dictámenes dispares y sintetizar observaciones con una versatilidad que, según algunos estudios, se aproxima a la del trabajo editorial humano (Kankanhalli, 2024).

Aunque las editoriales proclaman su compromiso con la ética y la transparencia, pocas explican de qué modo emplean la IA en sus procesos internos. El informe de COPE reconoce, de hecho, que la mayoría de los editores recurren a ella para el cribado, la detección de fraude o la sugerencia de revisores sin informar explícitamente a autores/as ni al público (Zhou y Soulière, 2025). Esta zona gris constituye lo que cabe definir como una alegalidad funcional, un espacio de práctica que no infringe normas explícitas porque esas normas aún no existen, pero operan al margen de cualquier mecanismo de escrutinio, trazabilidad o

rendición de cuentas. Se trata de una forma de anomia al servicio de intereses particulares, que no solo es aceptada y reproducida, sino que a menudo es deliberadamente buscada e incluso promovida. El discurso normativo apela a la transparencia y la honestidad intelectual mientras el sistema se sostiene sobre una arquitectura privativa que reproduce las mismas desigualdades que esas apelaciones decían querer corregir.

La retórica de la supervisión humana en todas las fases del proceso, habitualmente invocada bajo la fórmula del *human-in-the-loop*, rara vez se acompaña de mecanismos procesales concretos que permitan a los autores exigir responsabilidades en caso de conflicto. La evidencia empírica confirma que la transparencia mejora la calidad misma de la evaluación científica. Según el reciente estudio de Álvarez-García et al. (2026), los informes de revisión publicados en abierto presentan un contenido informativo notablemente superior al de los informes confidenciales y se concentran más en sugerencias constructivas. Organismos como la Association for Computing Machinery (ACM, 2023) y el CSIC (2026) coinciden en señalar que sin mecanismos claros de impugnación y reparación no puede garantizarse la fiabilidad de los sistemas algorítmicos (Makarem et al., 2023; Baeza-Yates, 2024). Sin embargo, los propios autores raramente ejercemos ese derecho más allá de la réplica inmediata al rechazo, una resignación práctica que termina normalizando la opacidad y erosionando la legitimidad del sistema desde dentro.

La literatura especializada lleva años documentando los sesgos de clase, etnia, lengua e institución que estructuran el funcionamiento de los algoritmos. En el ámbito editorial, estos sesgos no solo persisten sino que se amplifican. Farber (2024, p. 6) muestra que las revistas con recursos para implementar IA en la revisión reducen significativamente sus tiempos de decisión, pero al mismo tiempo aumentan la tasa de aceptación de artículos procedentes de instituciones de alto prestigio. El sesgo algorítmico y el sesgo estructural se retroalimentan. Los modelos aprenden de los corpus dominantes, reproducen sus jerarquías y penalizan los enfoques que se apartan de esos patrones (Mühlhoff, 2025). Los trabajos procedentes de regiones periféricas o de tradiciones críticas se enfrentan así a una doble invisibilidad: la de las métricas y la del algoritmo. Como señala Thelwall (2019), la automatización de la evaluación científica tiende a reforzar las asimetrías de capital académico y a erosionar la diversidad epistémica bajo la apariencia de neutralidad técnica.

La inteligencia artificial, incluso en su versión generativa, no es una entidad autónoma sino la sedimentación de decisiones humanas codificadas en sistemas cuya trazabilidad resulta con frecuencia opaca o directamente inaccesible. Los modelos aprenden de los corpus que les proporcionamos, reproducen nuestros sesgos y amplifican nuestras jerarquías. Como muestra el estudio de Ebadi et al. (2025), incluso los sistemas de recomendación de revisores/as tienden a favorecer a autores/as de instituciones anglófonas y a penalizar la diversidad lingüística. A ello se suma que ciertos patrones de alta frecuencia en los datos de preentrenamiento pueden distorsionar el aprendizaje del modelo, introduciendo sesgos que no siempre son detectables ni corregibles (Zevallos, Farrús y Bel, 2023). El sesgo de entrenamiento se traduce así en sesgo de legitimación. De ahí la importancia de garantizar que los datos editoriales (desde las bases de revisores/as hasta los repositorios de artículos) sean abiertos, plurales y representativos. La apertura de los datos no es un imperativo técnico, sino un principio democrático que se extiende también a la evaluación posterior a la publicación, donde los artículos pueden seguir siendo sometidos a revisión y valoración colectiva de manera continuada.

La expansión de la IA se ha producido, además, en un contexto de crisis estructural de la revisión por pares, marcada por la fatiga de revisores/as, la precarización del trabajo editorial y el incremento masivo de envíos. Hosseini y Horbach (2023) señalan que el volumen anual de manuscritos se ha duplicado en la última década, mientras que la disponibilidad de revisores/as cualificados ha crecido menos del veinte por ciento. En el ámbito de la sociología

española, la propia *Revista Española de Sociología* triplicó el número de manuscritos recibidos entre 2017 y 2023 (Cárdenas, 2023), una tendencia representativa del conjunto del ecosistema editorial en ciencias sociales. Este desajuste tiene consecuencias que van más allá de la logística editorial y erosiona el modelo de comunidad científica sobre el que se sostenía el sistema. En los primeros años del siglo, antes de la explosión del mercado editorial y de la proliferación de revistas cuya existencia responde más a la lógica de la inversión que a la del conocimiento, era la propia comunidad la que aportaba revisores con competencia suficiente y con un interés genuino en el estado del arte de su campo. En los últimos quince años esa dinámica se ha transformado radicalmente. Las revistas contactan a revisores con los que no tienen relación previa, publican en plazos cada vez más cortos e incentivan una competitividad que presiona a investigadores jóvenes a publicar mucho antes de haber consolidado su formación.

Ante este desajuste, la IA se presenta como una solución pragmática, un parche tecnológico para un problema sistémico. Lo que se gana en eficiencia se pierde en deliberación y pluralidad epistémica. La homogeneización de la evaluación tiende a desplazar la diversidad de criterios disciplinarios y, con ello, a empobrecer las condiciones mismas en que se produce el conocimiento. No es una hipótesis especulativa, pues el descenso de la calidad científica de las publicaciones ya se ha documentado en relación con el uso acrítico de la IA (Naddaf, 2025).

En este escenario, los grandes consorcios editoriales experimentan con IA bajo retóricas de innovación y sostenibilidad, pero con modelos cerrados y datos propietarios. Las revistas independientes y públicas (a menudo de ruta diamante) se encuentran, en cambio, entre la tentación de recurrir a la IA y la imposibilidad material de hacerlo en condiciones propias. Sin recursos para desarrollar modelos específicos y presionadas por criterios de indexación, oscilan entre la abstención y el rechazo de principio. Pero la brecha no es solo económica o infraestructural, es también disciplinar. Las ciencias experimentales, habituadas a la estandarización de procedimientos y al uso intensivo de herramientas computacionales, se han adaptado con relativa rapidez a la nueva centralidad de la IA. Las ciencias sociales y las humanidades, más dependientes del juicio interpretativo y del contexto, reaccionan con mayor cautela, o directamente con rechazo. La brecha tecnológica se convierte así en una brecha epistémica. La evidencia empírica muestra que las revistas de ciencias sociales presentan los estándares más elevados en la función deliberativa de la revisión por pares, por encima de las ciencias experimentales y de la salud (García-Costa et al., 2022). Su especificidad epistémica no es una limitación, sino su principal activo en este debate. Las revistas de estas disciplinas enfrentan un doble desafío: no quedar rezagadas tecnológicamente y, al mismo tiempo, no ceder su soberanía epistémica a modelos y lógicas ajenas a su especificidad metodológica e interpretativa.

El problema no es adoptar o rechazar la IA, sino en qué condiciones hacerlo. Rechazar la IA por principio equivale a renunciar a participar en el debate sobre su diseño, su ética y su función social. Una ética del no uso puede ser tan irresponsable como un uso acrítico, porque ambos extremos perpetúan la desigualdad. La tarea urgente para las revistas independientes no es ignorar la IA, sino construir marcos de gobernanza que la sometan a la deliberación colectiva y al control público. La respuesta no vendrá de los grandes consorcios ni de las corporaciones tecnológicas, sino de las comunidades científicas capaces de prefigurar alternativas.

La revisión por pares nació como un sistema humano de confianza, no como un mecanismo algorítmico de validación. Su crisis actual no se debe solo a la saturación de manuscritos, sino a la erosión de esa confianza. La IA puede agravar esa crisis si se utiliza como sustituto de la responsabilidad humana, pero también puede contribuir a restaurarla si se integra con criterios éticos y transparentes. El dilema no es técnico sino de gobernanza: quién define las reglas, quién audita los algoritmos, quién se beneficia de su uso. A partir de la distinción entre *regular*

(concebir a los agentes, humanos o no, como autónomos, capaces de cumplir o transgredir las normas) y *regimentar* (reducir su autonomía programando el cumplimiento en el propio diseño del sistema), Casanovas (2025) propone una estrategia híbrida que combina una arquitectura social (regulatoria) con una arquitectura técnica (regimentativa). El objetivo no es imponer una normativa externa sino integrar la ética en el diseño mismo de los sistemas, de modo que los valores de transparencia, trazabilidad y control humano sean propiedades constitutivas de las herramientas y no añadidos posteriores (Casanovas et al., 2025; Noriega et al., 2023).

Esta perspectiva conecta con el problema técnicamente conocido como alineación de valores (en inglés, *value alignment problem*), o cómo garantizar que los sistemas de IA actúen de acuerdo con los valores que las comunidades que los usan quieren preservar. La formulación más influyente de este problema sigue siendo la de Russell (2023), quien argumenta que los sistemas de IA no deben optimizar objetivos fijos sino operar con humildad epistémica respecto a las preferencias humanas y con capacidad de deferencia hacia el juicio humano, una idea que subyace a los marcos de gobernanza que aquí proponemos (Osman y Steels, 2024; Casanovas y Noriega, 2026).

Este horizonte de gobernanza encuentra correspondencia en marcos regulatorios como la Ley Europea de IA (Parlamento Europeo y Consejo de la Unión Europea, 2024) y un desarrollo concreto, aun en elaboración, en el modelo de cumplimiento multinivel propuesto en Gordo et al. (2025). Su principio central es que la transparencia, la trazabilidad y la supervisión humana deben ser propiedades constitutivas de las herramientas de IA, no añadidos posteriores. Más precisamente, sus autores proponen hablar de cumplimiento incorporado a través del diseño (en inglés, *compliance through design*): no se trata de programar el cumplimiento de reglas fijas, sino de diseñar sistemas lo suficientemente abiertos como para incorporar distintas interpretaciones de los valores que los fundamentan (Casanovas et al., 2025). Este modelo incluye seis niveles (criterios académicos compartidos, modelado y entrenamiento, integridad científica, salvaguardas éticas, gobernanza institucional y despliegue sostenible) que no constituyen una secuencia jerárquica sino un sistema de retroalimentaciones. Cada nivel incorpora mecanismos de trazabilidad y control humano que abarcan no solo los datos sino también las decisiones editoriales y las intervenciones del algoritmo: cada sugerencia automatizada queda documentada, cada error puede identificarse y corregirse, y cada decisión puede discutirse colectivamente. Aplicado a la revisión por pares, este enfoque implica que la sostenibilidad editorial no puede reducirse a la optimización de tiempos y procesos de gestión, sino que debe entenderse como una forma de justicia epistémica. Es un modelo de deliberación asistida, no de sustitución.

Un modelo con esas características puede transformar el ecosistema de las revistas de ruta diamante, al reducir la dependencia de plataformas comerciales cerradas y favorecer la construcción de infraestructuras abiertas e interoperables. Frente a la tendencia de las editoriales comerciales a convertir cada avance tecnológico en un servicio de pago, la alianza entre universidades, plataformas de software libre y comunidades editoriales puede articular un modelo alternativo de inteligencia editorial común, un espacio donde la IA opere como infraestructura compartida y no como producto propietario, y donde la deliberación colectiva prevalezca sobre la optimización privada. Desde esta perspectiva, la IA deja de ser una amenaza para convertirse en un instrumento de reequilibrio entre modelos de publicación con recursos desiguales. Alcanzar ese equilibrio, sin embargo, requiere voluntad política, coordinación institucional y una cultura de cooperación todavía débil en buena parte del ámbito académico.

Construir ese horizonte exige infraestructuras abiertas, auditables y gobernadas colectivamente, que garanticen a las comunidades científicas el control sobre los procesos de evaluación y legitimación del conocimiento. Frente al modelo de plataformas comerciales que concentran poder, el enfoque de cumplimiento multinivel se despliega mediante prácticas y tecnologías distribuidas, donde la IA actúa como soporte de deliberación y no como sustituto

de esta. No se trata de humanizar la IA, tratarla como si fuera una entidad externa a la que hay que dotar de valores, sino socializar su diseño, orientarla hacia los principios de cooperación, transparencia y rendición de cuentas que sustentan la ciencia abierta.

La cuestión de la gobernanza ya no puede delegarse. Las comunidades científicas deben recuperar capacidad de decisión sobre las herramientas con las que trabajan, y ello exige formación técnica, marcos institucionales de cooperación y voluntad política. Las directrices europeas recomiendan que los investigadores sean *en última instancia responsables* del contenido generado con apoyo de IA, pero esa responsabilidad individual es insuficiente sin estructuras colectivas de supervisión y participación. La ética de la agencia humana debe ampliarse hacia una ética de la comunidad. No basta con que cada investigador responda por su uso de la IA si el sistema en su conjunto no está sujeto a deliberación colectiva. En este punto, el papel de las revistas de ruta diamante es fundamental. Por su carácter público y su independencia económica, pueden convertirse en laboratorios de gobernanza distribuida, en espacios donde se construyan y consoliden formas alternativas de inteligencia editorial abierta.

La democratización de la IA en la edición científica no se logrará mediante declaraciones de principios, sino a través de la construcción cotidiana de estructuras transparentes y verificables. La supervisión debe ser real, los errores deben poder corregirse y las decisiones deben poder impugnarse. La integridad científica, en este nuevo contexto, no consiste en prohibir el uso de la IA, sino en garantizar que su intervención sea visible, contextualizada y sometida a deliberación colectiva. El conocimiento científico siempre ha dependido de mediaciones técnicas; lo que distingue a la IA es su capacidad de clasificar, de evaluar y, en última instancia, de decidir. La crítica sociológica debe desplazarse de la herramienta al sistema, del algoritmo al régimen de poder que lo sostiene.

La brecha de la IA en la edición científica no es un simple desequilibrio tecnológico. Es una cuestión de valores y de elección colectiva sobre qué modelo de ciencia queremos sostener. Cerrar esa brecha no significa renunciar a la tecnología, sino restituirla al control colectivo. Las revistas, los editores, los revisores, los autores y los lectores son actores de un mismo proceso de legitimación que solo puede sostenerse sobre la confianza y la transparencia.

Redefinir ese proceso de legitimación supone retomar una pregunta fundamental. Según recuerda Irfanullah (2023), la revisión por pares nació como un sistema humano de confianza, no como un mecanismo algorítmico de validación. Y la confianza no puede programarse sino construirse a través de la transparencia y la reciprocidad. Ante la consolidación de la IA en los procesos editoriales, la pregunta ya no es si debemos aceptarla, sino cómo vamos a gobernarla sin perder la dimensión pública y deliberativa del conocimiento. La respuesta exige comprender que la automatización no sustituye la responsabilidad y que la eficiencia no puede ser el único criterio de valor. La inteligencia artificial, si se somete a esos principios, puede convertirse en una aliada para recuperar el sentido público de la ciencia. Si se la abandona al mercado, será otro instrumento de concentración y exclusión.

REFERENCIAS

Álvarez-García, E., García-Costa, D., Squazzoni, F., Malički, M., Mehmani, B., y 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, 27 de junio). *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., y Noriega, P. (2026). Governance of Artificial Agency and AI Value Chains: A Few Remarks on Autonomy from a Legal and Ethical Approach. En D. López Castro, M. Cebral, y 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., y Lam, H. P (2025). Compliance, Regtech, and smart legal ecosystems: a methodology for legal governance validation. En 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., y 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., y 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. http://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., y Cester Mazarico, L. (2025, 4 de julio). *Aligning Generative AI with Scientific Responsibility: A Multi-Level Compliance Model for Peer Review.* [Documento de trabajo no publicado].

Hosseini, M., y 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, 29 de septiembre). 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., y 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, 29 de mayo). 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., y d'Inverno, M. (2023). Coordination, organizations, institutions, norms, and ethics for governance of multi-agent systems. En N. Fornara, J. Cheriyan, y 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., y 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.

Parlamento Europeo y Consejo de la Unión Europea (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., … y 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., y Bel, N. (2023). Frequency balanced datasets lead to better language models. En *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., y Soulière, M. (2025, 25 de agosto). *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/