

¿Seguirá siendo revisión por pares? Algoritmos, dictaminadores y la paradoja de Teseo

Raúl Marcó del Pont Lalli^{1*}

¹ Instituto de Geografía, Universidad Nacional Autónoma de México

* Dirección para correspondencia: edito@geografia.unam.mx

Cuenta un mito griego que Teseo, el héroe de Atenas, regresó victorioso en un barco que sus conciudadanos decidieron conservar como reliquia. Con el paso de los siglos, las tablas podridas fueron reemplazándose una por una. Cuando los filósofos repararon en lo ocurrido, se preguntaron si aquel barco seguía siendo el mismo. Y cuando la última tabla original hubo sido sustituida, la pregunta se volvió urgente: ¿en qué momento dejó de ser el barco de Teseo?

Vale la pena hacernos una pregunta análoga sobre la revisión por pares. Si sustituimos al revisor humano por un modelo de lenguaje —en parte o en su totalidad—, ¿seguimos hablando de revisión por pares? ¿O navegamos ya en otro barco, aunque le hayamos conservado el nombre?

Una historia más corta de lo que pensamos

La versión oficial atribuye la invención de la revisión por pares a Henry Oldenburg, secretario de la Royal Society de Londres, quien en 1665 compiló las primeras *Philosophical Transactions* —la primera revista científica del mundo— y

supuestamente consultó a expertos sobre lo que publicaba. La historia, difundida a partir de un célebre artículo de Harriet Zuckerman y Robert Merton en 1971, se repite desde entonces como hecho indiscutible. El problema es que no es cierto. Como ha demostrado la historiadora Melinda Baldwin, Oldenburg rara vez consultaba opiniones externas: decidía según su propio criterio y aceptaba tanto artículos experimentales como correspondencia personal.

El verdadero inventor del arbitraje escrito fue William Whewell, el polímata de Cambridge que también acuñó los términos "físico" y "científico". En 1831 propuso que la Royal Society comisionara informes escritos sobre los trabajos enviados y que se publicaran para fomentar el debate. La idea se adoptó a medias: se pidieron los informes, pero no se publicaron. Pronto los dictámenes se convirtieron en documentos confidenciales que los autores, durante décadas, ni siquiera recibían. Diversos estudios históricos han señalado que la revisión por pares no surgió originalmente solo como un mecanismo destinado a detectar fraude o garantizar la calidad de la literatura científica, sino también como una forma de proteger la reputación institucional de las sociedades científicas y administrar la circulación del conocimiento.

Durante todo el siglo XIX y buena parte del XX, el arbitraje externo fue una práctica opcional y desigual. Aileen Fyfe ha documentado cómo la Royal Society distribuyó el trabajo editorial entre secretarios, comités y árbitros, en un sistema pensado para diluir la responsabilidad. Revistas como *Nature* o *The Lancet* mantuvieron el control editorial en manos de sus directores hasta bien entrado el

siglo XX. Dos anécdotas fijan el contraste. En 1936, Einstein se indignó al descubrir que *Physical Review* había enviado su artículo sobre ondas gravitacionales a un revisor anónimo; no lo había autorizado, escribió al editor, y no volvió a someter un artículo a esa revista. En el otro extremo, Watson y Crick publicaron en 1953 su artículo sobre la doble hélice del ADN —acaso el más citado en la historia de *Nature*— sin revisión externa alguna. El sistema habría podido rechazar o retrasar ambos trabajos. Sin embargo, la revisión por pares también ha funcionado históricamente como un mecanismo capaz de contener errores metodológicos, detectar inconsistencias o frenar publicaciones deficientes. La cuestión no es negar esa función, por demás relevante, sino recordar que el sistema ha sido siempre más heterogéneo, conflictivo e imperfecto de lo que suelen sugerir sus defensas institucionales.

La Guerra Fría y el nacimiento de un dogma

Si la revisión por pares existe hoy como procedimiento universal y símbolo de legitimidad científica, la razón no está en el siglo XVII sino en los años setenta del XX, en plena Guerra Fría estadounidense. En 1975, el senador William Proxmire lanzó su célebre Premio Golden Fleece —una insignia mensual de deshonra— a los proyectos que consideraba el peor uso del dinero de los contribuyentes. Los dos primeros se otorgaron a proyectos de la Fundación Nacional para la Ciencia (NSF, por sus siglas en inglés): un estudio sobre la atracción interpersonal y otro sobre por

qué las ratas y los monos aprietan los dientes bajo estrés. Un precursor encorsetado de los IG Nobel.

La controversia derivó en unas audiencias del Congreso en julio de 1975. Y ahí ocurrió algo decisivo: todas las partes —críticos y defensores, legisladores y científicos— coincidieron en que la revisión por pares era el mecanismo fundamental para garantizar la calidad de la ciencia. Quienes atacaban a la NSF no atacaban el *dictamen por pares*: sostenían que la NSF no lo aplicaba bien. Por ello, la Fundación reforzó entonces el papel de los árbitros externos, comenzó a compartir los dictámenes con los solicitantes y creó una oficina de auditoría. En pocos años, la revisión por pares pasó de ser uno entre varios métodos editoriales aceptables a ser el único mecanismo legítimo de validación científica. El término mismo tiene una historia reveladora: Baldwin ha rastreado que no surgió en las revistas sino en los comités médicos de Medicare y Medicaid. Lo que llamamos revisión por pares tiene apenas medio siglo como concepto unificado.

El guardián y sus grietas

Con el paso de los años, la revisión por pares acumuló expectativas muy por encima de lo que su diseño original podía sostener: detectar fraude, garantizar la reproducibilidad, distribuir el reconocimiento con equidad, distinguir la buena ciencia de la mala con precisión milimétrica. No ha podido hacer ninguna de esas cosas de manera consistente. Un metaanálisis de 45 estudios citado por Aczél y

colaboradores en las *Proceedings of the National Academy of Sciences of the United States of America* (PNAS) de 2025 encontró que la correlación promedio entre las evaluaciones de dos revisores del mismo manuscrito es apenas de 0.34: un nivel de acuerdo débil, considerablemente menor de lo que cabría esperar de un mecanismo presentado como árbitro robusto de la calidad científica.

Las consecuencias son conocidas. En psicología, el 18 % de los resultados estadísticos publicados contiene errores de reporte. La tasa de retractaciones crece año con año. Abundan los ejemplos de trabajos rechazados que más tarde se volvieron clásicos o merecieron premios Nobel: la investigación de Katalin Karikó sobre el mRNA fue ignorada por el sistema que supuestamente garantiza la calidad. Los sesgos tampoco son menores: la afiliación institucional influye en las aceptaciones, y el doble ciego, diseñado para mitigarlo, no lo elimina —el campo, el estilo y las referencias pueden revelar la identidad.

El problema estructural más profundo es que la revisión por pares es trabajo no remunerado que no cuenta para la promoción académica. Aczél *et al.* calcularon que el tiempo dedicado colectivamente a revisar representa una donación anual valorada en miles de millones de dólares, extraída gratuitamente por un sistema editorial cuyas versiones comerciales generan márgenes superiores al 30 %. Según Publons, el número de invitaciones necesarias para conseguir que un solo revisor acepte un manuscrito aumentó de 1.9 en 2013 a 2.4 en 2018. En este contexto llegó la inteligencia artificial.

La IA en el barco

Una revisión exploratoria publicada en abril de 2026 en el *International Journal of Medical Informatics*, firmada por Nabavi y colaboradores, sintetizó 189 fuentes sobre el uso de la IA en la revisión por pares. El panorama es inequívoco: los modelos de lenguaje de gran escala ya están integrados en los flujos editoriales de muchas revistas, y la velocidad de adopción supera a los organismos reguladores. Para agosto de 2025, las políticas seguían siendo fragmentadas y contradictorias, y el uso de IA generativa estaba a menudo más restringido para los revisores que para los propios autores: una asimetría reveladora. Los usos van desde la selección inicial hasta la generación completa de dictámenes, con beneficios reales en eficiencia, estandarización y ampliación del conjunto de revisores.

Pero los problemas son igualmente nítidos. Los sistemas actuales carecen de razonamiento disciplinar específico y de juicio ético para la evaluación autónoma. El sesgo algorítmico es real: los modelos entrenados con literatura existente tienden a favorecer las instituciones de élite, los formatos convencionales y los autores con mayor visibilidad previa, reproduciendo exactamente los sesgos que el doble ciego fue diseñado para combatir. La confidencialidad es otro punto crítico: cuando un revisor carga un manuscrito en una herramienta de IA de terceros, puede estar violando el acuerdo implícito que sostiene el proceso. Y hay algo más difícil de medir: la homogeneización de la voz académica. Si todos los dictámenes pasan por los mismos modelos, la evaluación podría converger hacia un estilo uniforme que

privilegie cierto tipo de argumentos, cierta estructura, cierta gramática del conocimiento. La monocultura intelectual es uno de los peligros más serios del sistema actual, y la IA podría intensificarla. Aunque también es posible, al menos hipotéticamente, que sistemas híbridos bien diseñados amplíen el acceso a voces evaluativas hoy excluidas por barreras geográficas o de idioma.

Leer un texto, juzgar un campo

Las comparaciones empíricas entre dictámenes humanos y generados por IA revelan algo fundamental. Los modelos de lenguaje reproducen con notable fidelidad la forma del dictamen académico: estructura retórica, evaluación general, observaciones por secciones, recomendación editorial. Identifican problemas de redacción, inconsistencias internas, debilidades organizativas. Son lectores competentes del texto. Pero cuando se trata de juzgar si un artículo hace una contribución disciplinaria genuina —si no solo describe un fenómeno, sino que lo interpreta de una manera que avanza el campo—, los modelos muestran limitaciones significativas. Los revisores humanos sitúan el manuscrito en una conversación: ¿qué dice este trabajo que no se haya dicho antes? ¿A qué debate responde? Los modelos, en cambio, evalúan la coherencia interna: si el argumento es consistente, si la literatura citada parece adecuada, si la estructura resulta plausible.

Podría formularse así: la IA es una buena lectora del texto; los revisores humanos son mejores jueces del conocimiento.

La distinción se vuelve aún más relevante a la luz de lo que se ha identificado como la función más profunda de la revisión por pares: no solo filtrar la ciencia deficiente, sino crear comunidad científica. Los revisores y revisados tienden a acercarse en sus redes incluso bajo doble ciego, y contribuyen juntos a estandarizar los criterios evaluativos de su campo. Revisar es también aprender y construir disciplina. Un algoritmo puede imitar el producto —el texto del dictamen—, pero no puede participar en él. La revisión por pares no es solo corrección de estilo ni verificación lógica: es un mecanismo de reproducción del conocimiento disciplinario. A través de ella, una comunidad de especialistas decide qué cuenta como contribución, qué preguntas merecen atención, qué métodos son legítimos. Es, en el fondo, un acto político y epistemológico, no solo editorial.

El robot que aprendió a ser humano

Hay un episodio de *Doctor Who*, titulado “Respira hondo” (2014), en el que el doctor se enfrenta a un androide que ha sobrevivido durante siglos reemplazando sus piezas desgastadas con órganos humanos. El robot lleva tanto tiempo haciéndolo que ya no recuerda si alguna vez fue completamente mecánico. La pregunta del robot y la del barco de Teseo son la misma, y ambas se aplican a lo que ocurre con la revisión por pares. Primero llegaron las plataformas de gestión de manuscritos.

Luego los detectores de plagio. Después los algoritmos de recomendación de revisores. Ahora los modelos de lenguaje reemplazan partes del juicio crítico. Cada sustitución se justifica con el mismo argumento: eficiencia, velocidad, escalabilidad. Y cada vez alguien señala que el barco sigue navegando.

Pero hay una diferencia crucial entre el barco de Teseo y el dictamen académico. Las tablas del barco eran intercambiables porque cumplían la misma función. Un revisor humano y un modelo de lenguaje no son intercambiables si hacen cosas distintas. Y las hacen. El revisor humano no solo lee: *pertenece a una comunidad* cuyos valores, debates y tradiciones están en juego en cada dictamen que emite. El modelo de lenguaje no pertenece a ninguna comunidad disciplinaria. Sin embargo, más que una afirmación empírica estricta, esta idea, debemos reconocer, remite a una vieja y no saldada discusión en sociología y filosofía de la ciencia sobre la relación entre experiencia disciplinaria, pertenencia comunitaria y legitimidad del juicio experto.

Aczél y colaboradores dedican además un apartado —provocadoramente titulado “¿Abandonar el dictamen por pares?”— a tomar en serio que el sistema se elimine en lugar de reformarse. El argumento histórico es poderoso: la ciencia se hizo con notable éxito durante siglos sin revisión por pares sistemática. El artículo de Watson y Crick se publicó sin ella; los *Annus mirabilis* de Einstein en 1905, tampoco. Y el dictamen por pares tiende a crear monocultura intelectual: juzga los trabajos contra el trasfondo del conocimiento ya validado, cuando la mayoría de los avances paradigmáticos ocurrieron *a pesar* del consenso vigente, no gracias a él.

No proponemos eliminar la revisión por pares. Pero tomarse en serio esa posibilidad obliga a precisar qué defendemos cuando la defendemos. ¿El proceso? ¿El resultado? ¿La comunidad que crea? ¿El cerrojo que mantiene?

El barco que vale la pena conservar

Nada de esto significa que la inteligencia artificial no tenga lugar en los procesos editoriales. Significa que ese lugar debe definirse con precisión y honestidad. Hay tareas en las que los modelos de lenguaje ofrecen un apoyo genuino: detección de inconsistencias formales, revisión de estilo, selección preliminar de manuscritos, apoyo a editores de revistas pequeñas. El problema no es la herramienta; es la confusión entre la herramienta y el proceso que pretende apoyar. También merecen considerarse reformas del proceso humano, algo que ha generado largas y agrias controversias. La revisión abierta tampoco está exenta de problemas. Diversos estudios han señalado que la visibilidad de la identidad de los evaluadores puede favorecer formas de autocontención crítica, especialmente entre investigadores jóvenes o en comunidades académicas pequeñas.

Lo que no puede hacerse —o no sin consecuencias graves para la validación del conocimiento— es delegar en un algoritmo la pregunta central: ¿este trabajo contribuye a la comprensión que tenemos del mundo? La evidencia sobre la función comunitaria del arbitraje sugiere por qué: los revisores no solo evalúan textos, sino que, al hacerlo, construyen los criterios mismos con los que su campo define una

contribución válida. Esa pregunta, por tanto, solo puede responderla con autoridad alguien que habita ese mundo, que conoce sus debates, que tiene algo en juego en su respuesta, y que participa, con ella, en la reproducción del campo al que pertenece.

La revisión por pares no tiene un largo abolengo. Es una práctica reciente, consolidada en los años setenta del siglo XX bajo presiones políticas y presupuestarias, no epistemológicas. Ha cambiado antes y puede cambiar de nuevo. Pero la dirección del cambio importa, y en ese debate los editores, los revisores y los investigadores de países como México tienen más voz de la que suelen ejercer. El barco de Teseo puede navegar con tablas nuevas, e incluso incorporar materiales que Teseo nunca imaginó. Lo que no puede hacer es navegar sin nadie que sepa a dónde va.

El dramaturgo suizo Max Frisch definió la tecnología como “el truco que consiste en organizar el mundo de modo que no tengamos que experimentarlo”. La cita resuena hoy con una precisión que incomoda. Si cedemos a la tentación de delegar en la inteligencia artificial la revisión por pares —ese momento en que una comunidad científica se mira al espejo y decide qué merece existir como conocimiento—, corremos el riesgo de organizar el mundo académico de tal manera que no tengamos que habitarlo. Que no tengamos que leer, ni juzgar, ni equivocarnos, ni aprender de la equivocación. La pregunta no es si la IA puede hacer esto. La pregunta es si queremos un sistema de producción de conocimiento en el que nadie haya tenido que experimentar nada.

Nota

Para la versión final de este escrito empleé Claude (Anthropic) con una función estrictamente auxiliar: ajuste de la fluidez narrativa. Las ideas, los argumentos, la estructura conceptual, la selección de fuentes y todas las afirmaciones sustantivas son de mi autoría; ningún contenido fue generado íntegramente por inteligencia artificial.

Referencias

Aczél B, Barwich A-S, Diekman AB and Ioannidis JPA (2025). The present and future of peer review: Ideas, interventions, and evidence. *Proceedings of the National Academy of Sciences* 122(5):e2401232121.

Baldwin M (2017). In referees we trust? *Physics Today* 70(2):44–49.

Baldwin M (2018). Scientific Autonomy, Public Accountability, and the Rise of “Peer Review” in the Cold War United States. *Isis* 109(3):538–558.

Fyfe A (2020). Editors, referees, and committees: Distributing editorial work at the Royal Society journals in the late 19th and 20th centuries. *Centaurus* 62:125–140.

Nabavi A, Safari F, Shmoury AH *et al.* (2026). Artificial intelligence in scholarly peer review: a scoping review of applications, risks, and governance challenges. *International Journal of Medical Informatics* 214:106418.

Publons (2018). *Global state of peer review 2018*. Publons.

Zuckerman H and Merton RK (1971). Patterns of evaluation in science: Institutionalization, structure, and functions of the referee system. *Minerva* 9(1):66–100.