Masferrer Investiga
Revista Científica de la Universidad Salvadoreña Alberto Masferrer - USAM

Revista Masferrer Investiga, Vol. 15, N° 3, [pp. 43-58] Septiembre, 2025
ISSN: 2079-9020 | e-ISSN: 2079-9039
DOI: https://doi.org/10.65880/ba5myc15

IA responsable en el ciclo de la investigación: revisión de alcance PRISMA-ScR (2024–2025)

Responsible AI in the research lifecycle: a PRISMA-ScR scoping review (2024–2025)

Recibido: abril 2025
Aceptado: Mayo 2025

Como citar este articulo (APA)
Oviedo-Argueta, A., Vásquez, B. P. A., & Pérez de Martínez, J. B. (2025). IA responsable en el ciclo de la investigación: revisión de alcance PRISMA-ScR (2024–2025). Masferrer Investiga, 15(3), 41-57. https://doi.org/10.65880/ba5myc15

Alberto Oviedo-Argueta
Universidad Salvadoreña Alberto Masferrer (El Salvador)
https://orcid.org/0009-0000-7474-2072
alberto.oviedo01@liveusam.edu.sv

Bruno Paúl Alexander Vásquez
Universidad Salvadoreña Alberto Masferrer (El Salvador)
https://orcid.org/0009-0006-5111-9230
bruno.vasquez01@liveusam.edu.sv

Johanna Beatriz Pérez de Martínez
Universidad Salvadoreña Alberto Masferrer (El Salvador)
https://orcid.org/0000-0002-4984-324X
johana.perez01@liveusam.edu.sv

Resumen

La irrupción de la IA generativa está reconfigurando el trabajo académico, equilibrando productividad e integridad. Esta revisión de alcance PRISMA-ScR, mapea la integración responsable de la IA a lo largo del ciclo de la investigación, búsqueda, escritura, revisión por pares, reporte y gobernanza editorial, y sintetiza prácticas aplicables en entornos universitarios y editoriales. La búsqueda en Scopus, limitada a inglés, acceso abierto y versión final entre 2024 y 2025, identificó 79 registros; tras el cribado temático se incluyeron 15 documentos que abarcan políticas y posicionamientos editoriales, estudios empíricos y metainvestigación, guías de reporte y propuestas metodológicas. El análisis mostró convergencia en principios clave: la IA no puede figurar como autora; la divulgación debe ser explícita e incluir herramienta, versión, propósito y, cuando corresponda, prompts; y los detectores presentan limitaciones y sesgos, por lo que no deben usarse como prueba única en decisiones editoriales. Las políticas evolucionan hacia esquemas escalonados con responsabilidad humana y resguardo de la confidencialidad, especialmente en revisión por pares, donde se desaconseja cargar manuscritos inéditos en servicios públicos y se acotan tareas asistibles por modelos bajo supervisión. Emergen listas de verificación y estándares de reporte que exigen identificar modelo y versión, semillas, prompts y criterios de evaluación, además de flujos con control de versiones para auditar contribuciones. Se documentan riesgos metodológicos, como la inestabilidad por repetición de prompts, lo que exige reproducibilidad y trazabilidad efectiva. Proponemos una arquitectura mínima de uso responsable que refuerza integridad, reproducibilidad y confianza en la comunicación científica.

Palabras clave: Inteligencia Artificial Generativa; Modelo De Lenguaje; Revisión Por Pares; Integridad Científica; Reproducibilidad.

Revista Masferrer Investiga, Vol. 15, N° 3, [pp. 43-58] Septiembre, 2025
ISSN: 2079-9020 | e-ISSN: 2079-9039
DOI: https://doi.org/10.65880/ba5myc15

Abstract

Generative artificial intelligence is reshaping scholarly work by expanding productivity while demanding rigorous safeguards for research integrity. This scoping review, aligned with PRISMA ScR, maps responsible integration of AI across the research lifecycle, including search, writing, peer review, reporting, and editorial governance, and synthesizes practices for universities and journals. A structured Scopus search limited to English open access and final status during 2024 and 2025 identified seventy-nine records. After screening, fifteen documents were included, spanning editorial policies and statements, empirical and meta-research, reporting guidance, and methodological proposals. Analysis revealed convergence on core principles. Artificial intelligence cannot be credited as an author. Disclosure must be explicit and granular, specifying tool, model version, intended purpose, and prompts when relevant. Given known limitations and biases, AI detectors should not constitute sole evidence in editorial decisions. Policies are evolving toward tiered use with human accountability and confidentiality safeguards. In peer review, uploading unpublished manuscripts to public services is discouraged, and permitted tasks for language models are narrowly defined under supervision. Reporting checklists clearly require identification of model and version, seeds, prompts, and evaluation criteria. Version controlled workflows enable audit trails of AI and human contributions. Methodological risks, including instability from prompt repetition, underscore the need for reproducibility and robust traceability. We propose a minimal architecture for responsible adoption that implements disclosure matrices, human accountability, a limited role for detectors, supervised peer review protocols, reporting checklists, and transparent change logs. This architecture strengthens integrity, reproducibility, and trust across scholarly communication.

Keywords: Generative Artificial Intelligence; Large Language Model; Peer Review; Research Integrity; Reproducibility.

Introducción

La aceleración reciente en el desarrollo y adopción de sistemas de inteligencia artificial está transformando de forma transversal el ciclo de la investigación científica: desde la formulación de preguntas y la búsqueda de literatura, hasta la recolección/curación de datos, el análisis, la redacción y la evaluación por pares. Este despliegue simultáneo de modelos generativos, asistentes de revisión, motores de descubrimiento y herramientas de análisis plantea oportunidades inéditas de eficiencia y alcance, pero también riesgos concretos que comprometen la validez, la equidad y la reproducibilidad del conocimiento. En este contexto, la noción de IA responsable, que integra principios de transparencia, explicabilidad, privacidad, seguridad, justicia, rendición de cuentas e inclusión, ha pasado de ser un ideal normativo a una condición operativa para salvaguardar la integridad científica (Akinrinola et al., 2024; Carter et al., 2019; Istasy et al., 2021).

A pesar del auge de marcos y guías de "IA confiable", la evidencia empírica sobre cómo esos principios se aplican, o fallan, en las tareas cotidianas de investigación sigue fragmentada. Existen recomendaciones generales y casos de uso parciales, pero falta una síntesis que conecte los riesgos, salvaguardas y buenas prácticas etapa por etapa del ciclo de la investigación. El resultado es una brecha entre la formulación

ISSN: 2079-9020 | e-ISSN: 2079-9039
DOI: https://doi.org/10.65880/ba5myc15

ética de alto nivel y las decisiones prácticas que toman investigadores, comités editoriales, revisores, bibliotecas/CRAI y agencias financiadoras al implementar o regular herramientas de IA (Li et a., 2023). Para atender esta brecha, realizamos una revisión de alcance (PRISMA-ScR) centrada en el período 2024–2025, con el objetivo de mapear sistemáticamente la literatura reciente sobre IA responsable a lo largo del ciclo de la investigación, identificar patrones, vacíos y tensiones, y proponer una taxonomía operativa que vincule tareas, riesgos y salvaguardas. Guiaron nuestro trabajo las siguientes preguntas:

- ¿Qué riesgos y fallos de la IA emergen de forma diferencial en cada etapa del ciclo (búsqueda, datos, análisis, escritura, evaluación, difusión y preservación)?
- ¿Qué salvaguardas técnicas, organizativas y procedimentales han mostrado mayor promesa para mitigar esos riesgos sin sacrificar productividad ni rigor?
- ¿Qué vacíos de evidencia persisten (métricas, auditorías, reportes de impacto, gobernanza), y dónde se concentran?
- ¿Qué recomendaciones accionables se pueden ofrecer a actores clave (investigadores, CRAI/bibliotecas, editores, financiadores, comités de ética) para operacionalizar la IA responsable con trazabilidad y auditoría?

Adoptamos el enfoque PRISMA-ScR por su idoneidad para campos emergentes y heterogéneos, priorizando estudios publicados entre 2024 y 2025 que describen aplicaciones, riesgos, marcos, auditorías o intervenciones sobre IA en procesos de investigación (Polanco & Mayorga, 2025). Tras la depuración, 15 artículos cumplieron criterios de elegibilidad y fueron analizados mediante extracción temática y síntesis narrativa. Esta selección cubre disciplinas y escenarios diversos, lo que permite observar regularidades y variaciones contextuales sin imponer una homogeneidad artificial (Teodoro et al., 2025).

Nuestro trabajo aporta cuatro contribuciones principales. Primero, proponemos una taxonomía etapa-tarea que alinea principios de IA responsable con riesgos específicos y controles verificables (por ejemplo, trazabilidad de prompts y datasets; pruebas de robustez y sesgo; políticas de autoría y divulgación de uso de IA; salvaguardas de privacidad; estándares de citación y verificación de fuentes). Segundo, sintetizamos prácticas prometedoras, técnicas y organizativas, que muestran factibilidad y valor añadido en entornos reales. Tercero, identificamos vacíos críticos: métricas comparables de impacto, evidencia sobre efectos en revisión por pares, costos de cumplimiento en grupos con menos recursos, e interoperabilidad de reportes entre instituciones (Echeverría et al., 2023; Hernández et al., 2023; Mayorga et al., 2023). Cuarto, derivamos un conjunto de recomendaciones operativas orientadas a decisiones, con niveles de madurez y responsables claros, para facilitar su adopción incremental en universidades y revistas (Lu et al., 2024). El alcance de esta revisión es deliberadamente pragmático: no pretende dirimir debates filosóficos sobre la naturaleza de la inteligencia o la autonomía de los sistemas, sino traducir principios en controles implementables y en evidencias que permitan evaluar beneficios y riesgos con criterios compartidos. Consideramos que este enfoque es especialmente útil para actores que deben balancear innovación, cumplimiento y confianza pública, incluidos los CRAI que hoy articulan servicios, formación, curación de datos y soporte metodológico.

Revista Masferrer Investiga, Vol. 15, N° 3, [pp. 43-58] Septiembre, 2025
ISSN: 2079-9020 | e-ISSN: 2079-9039
DOI: https://doi.org/10.65880/ba5myc15

Materiales y método

El estudio adoptó una revisión de alcance para examinar de forma comprensiva la integración responsable de la IA generativa a lo largo del ciclo de la investigación. Esta elección metodológica responde al carácter emergente del campo y a la dispersión de la evidencia disponible, lo que hace pertinente un mapeo amplio antes que una síntesis evaluativa. La revisión se condujo conforme a la extensión PRISMA-ScR, que establece criterios mínimos y etapas secuenciadas para planificar, seleccionar, extraer y sintetizar de manera transparente.

Etapa 1. Planificación

Se delimitaron preguntas orientadas a: i) identificar riesgos y salvaguardas reportados por etapa del ciclo de la investigación; ii) caracterizar políticas, guías y estándares de uso y reporte de IA en publicación científica; y iii) localizar vacíos metodológicos y de gobernanza. Se diseñó una estrategia de búsqueda basada en términos controlados y sinónimos combinando "generative AI", "large language model", "LLM", "ChatGPT" con "research lifecycle", "scholarly publishing", "research integrity", "peer review", "authorship", "disclosure", "editorial policy", "guideline", "checklist", "PRISMA", "transparency". La búsqueda se ejecutó en Scopus por su cobertura multidisciplinaria y riqueza de metadatos. Para asegurar estabilidad editorial, se limitaron los resultados a documentos en inglés, acceso abierto, versión final, tipo artículo, publicados en 2024–2025. Se definieron exclusiones para trabajos técnico-clínicos sin vínculo con integridad o publicación, preprints, editoriales breves sin contenido normativo/metodológico, duplicados y documentos fuera de los filtros lingüísticos y de acceso (ver Tabla 1).

Tabla 1.

Criterios de inclusión y exclusión

Indicadores	Criterios de inclusión	Criterios de exclusión
Metodología / diseño	Artículos con enfoque cualitativo, cuantitativo o mixto, así como revisiones narrativas, de alcance o sistemáticas que analicen el uso responsable de IA en procesos de investigación, publicación o evaluación científica.	Estudios experimentales, preexperimentales o cuasiexperimentales enfocados en el desarrollo técnico de modelos de IA sin relación con la integridad, la ética o la publicación científica.
Fuente de publicación	Artículos publicados en revistas científicas revisadas por pares, capítulos de libro o actas de congreso con revisión editorial y relevancia académica.	Libros completos, informes institucionales, editoriales, comentarios de opinión, blogs, notas de prensa o cartas al editor sin revisión por pares.
Temática / sujeto	Artículos que aborden de forma explícita el uso de IA generativa o modelos de lenguaje (LLM) en el ciclo de la investigación, incluyendo escritura científica, revisión por pares, políticas editoriales, divulgación del uso de IA, reproducibilidad y estándares de reporte.	Trabajos sobre IA en contextos técnicos, clínicos o industriales sin relación con publicación científica, integridad académica o gobernanza editorial.
Periodo temporal considerado	Publicaciones comprendidas entre 2024 y 2025, priorizando artículos con versión final y acceso abierto publicados entre 2024 y 2025.	Publicaciones anteriores a 2024 o fuera del rango temporal establecido.

Idioma	Artículos publicados originalmente en inglés y disponibles en acceso abierto.	Artículos en otros idiomas o sin traducción disponible al inglés.
Cadena de búsqueda (03/11/2025)	(TITLE-ABS-KEY ("generative AI" OR "large language model*" OR LLM OR ChatGPT) AND TITLE-ABS-KEY (LLM OR "large language model*" OR ChatGPT) AND TITLE-ABS-KEY ("research lifecycle" OR "scholarly publishing" OR "research integrity" OR "peer review" OR authorship OR disclosure OR "scientific writing") AND TITLE-ABS-KEY (policy OR guideline OR "editorial policy" OR "author guidelines" OR governance OR transparency OR PRISMA OR "best practice*")) AND PUBYEAR > 2024 AND PUBYEAR < 2025 AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (PUBSTAGE , "final"))	

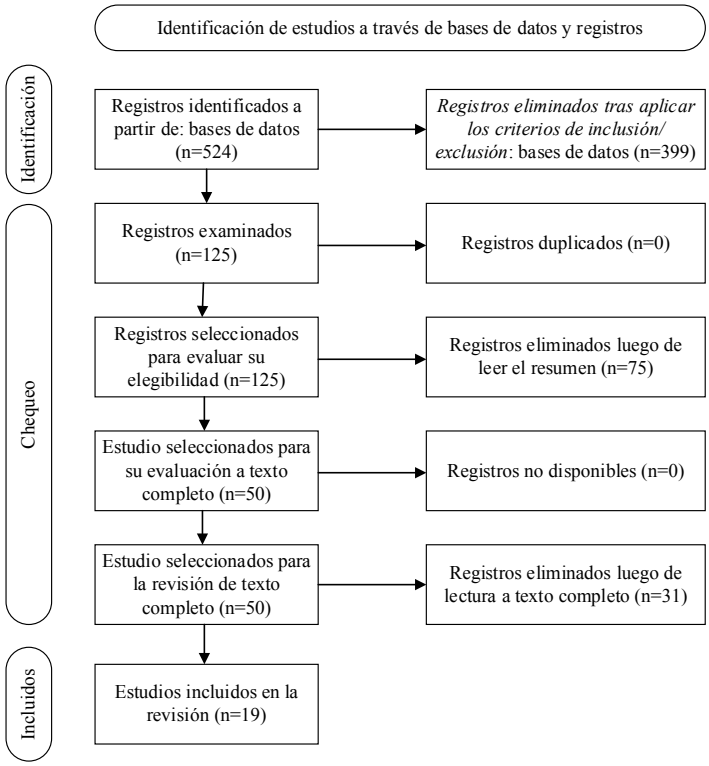
Etapa 2. Selección

La consulta recuperó 79 registro. Se realizó un cribado por título y resumen para pertinencia temática y, posteriormente, lectura a texto completo de los candidatos elegibles. La selección aplicó de forma consistente los criterios predefinidos y resolvió discrepancias por consenso. El conjunto final quedó constituido por 15 artículos que abordaban, con distinto énfasis, políticas editoriales, estudios empíricos y meta investigación, guías de reporte y propuestas metodológicas relacionadas con IA generativa en la comunicación científica (ver Figura 1, Tabla 2).

Revista Masferrer Investiga, Vol. 15, N° 3, [pp. 43-58] Septiembre, 2025
ISSN: 2079-9020 | e-ISSN: 2079-9039
DOI: https://doi.org/10.65880/ba5myc15

Figura 1.

Diagrama de flujo PRISMA que representa gráficamente el proceso de selección.



Etapa 3. Extracción de datos

Se desarrolló una plantilla estandarizada para registrar metadatos (título, revista, año, ámbito), tipo de contribución (política/posición, empírico/metainvestigación, guía/estándar, propuesta metodológica), etapa del ciclo impactada (búsqueda y síntesis, escritura, revisión por pares, reporte/comunicación, gobernanza editorial), requisitos de divulgación del uso de IA (herramienta, modelo/versión, propósito, prompts/semillas cuando procedía), postura sobre detectores y sus limitaciones, salvaguardas en revisión por pares (confidencialidad, supervisión humana, tareas permitidas y prohibiciones), estándares de reporte y exigencias mínimas, elementos de reproducibilidad y trazabilidad (control de versiones y auditoría de cambios), riesgos metodológicos identificados y recomendaciones operativas. La extracción fue pilotada en un subconjunto y verificada por un segundo revisor para asegurar consistencia. Dado el propósito exploratorio, no se aplicó una herramienta formal de riesgo de sesgo; sin embargo, se contextualizó la fuerza de la evidencia según la naturaleza de cada fuente y sus limitaciones declaradas.

Etapa 4. Síntesis

Se efectuó una síntesis narrativa temática, organizando los hallazgos en ejes previamente definidos y refinados inductivamente: gobernanza y divulgación del uso de IA, IA en revisión por pares, detección y su rol acotado, estándares de reporte y listas de verificación, y reproducibilidad con trazabilidad editorial. A partir de estos ejes se construyó un mapa etapa, riesgo, salvaguarda y se derivó una arquitectura mínima de adopción responsable con controles verificables orientados a universidades y revistas.

ISSN: 2079-9020 | e-ISSN: 2079-9039
DOI: https://doi.org/10.65880/ba5myc15

Resultados

La revisión integró quince estudios publicados entre 2024 y 2025, todos de acceso abierto, revisados por pares y centrados en el uso de inteligencia artificial generativa (IAg) o modelos de lenguaje de gran tamaño (LLM) en procesos de investigación y comunicación científica. Los artículos provienen de campos diversos, medicina, ciencias sociales, educación, ética, odontología y edición científica, lo que permite una visión transversal de los retos y convergencias que caracterizan la adopción responsable de la IA en el ciclo de la investigación. A partir del análisis temático se identificaron cuatro ejes principales: (1) gobernanza y divulgación del uso de IA; (2) IA en la revisión por pares; (3) detección y limitaciones de los verificadores automáticos; y (4) estándares de reporte y trazabilidad en la comunicación científica.

Gobernanza y divulgación del uso de IA

La mayoría de los estudios enfatiza la necesidad de establecer políticas editoriales claras que definan el alcance, los límites y la responsabilidad humana en el uso de IA durante la escritura y revisión científica. Yoo (2025) y Vasconcelos y Marušić (2025) coinciden en que la IA no puede figurar como autora, y su empleo debe revelarse explícitamente indicando la herramienta, modelo, versión y propósito. Se observa una transición desde declaraciones genéricas hacia esquemas de uso escalonado, en los que la responsabilidad recae en el autor principal y las instituciones patrocinadoras. Además, Lendvai (2025) muestra un aumento exponencial de publicaciones que incorporan ChatGPT o LLM, pero con reportes dispares de transparencia, lo que evidencia una falta de normalización en los formatos de divulgación. Los estudios proponen implementar matrices de revelación y registros auditables de prompts y versiones de modelo como mecanismos de trazabilidad y rendición de cuentas.

IA en la revisión por pares

Cinco artículos (Eggmann et al., 2025; Ebadi et al., 2025; Marrella et al., 2025; Mollaki, 2024; Lee et al., 2025) abordan directamente la integración de LLM en el proceso de revisión. Las evidencias convergen en que el apoyo de IA puede mejorar la consistencia lingüística y la detección de errores formales, pero conlleva riesgos de confidencialidad, sesgo y dependencia cognitiva. Eggmann et al. (2025) identifican el uso no declarado de IA en revisores de revistas odontológicas, mientras que Ebadi et al. (2025) evidencian posturas divididas: los revisores valoran la eficiencia pero temen comprometer la imparcialidad y la originalidad de los juicios. Marrella et al. (2025) comparan evaluaciones humanas y generadas por IA, concluyendo que los modelos pueden complementar, pero nunca sustituir, la evaluación experta. En conjunto, los estudios recomiendan protocolos con supervisión humana obligatoria, restricciones de carga de manuscritos inéditos a servicios públicos y delimitación precisa de tareas permitidas.

Detección y limitaciones de los verificadores automáticos

Erol et al. (2025) demuestran que los detectores de texto generados por IA presentan una alta tasa de falsos positivos y resultados inconsistentes entre plataformas, por lo que su uso como evidencia única en decisiones editoriales es inapropiado. La confiabilidad de estas herramientas depende del idioma, la longitud del texto y la versión del modelo base, factores que aún carecen de estandarización. En consecuencia, varios artículos (Yoo 2025; Teixeira da Silva, 2024; Mollaki, 2024) abogan por limitar los

detectores a funciones de apoyo técnico y no sancionador, complementándolos con auditorías humanas y revisión cruzada.

Estándares de reporte y trazabilidad

Los trabajos de Huo et al. (2025) y Al-Qudimat et al. (2025) introducen avances en guías de reporte específicas, como el CHART Statement, que exige detallar herramienta, versión, propósito, configuración y supervisión humana. Estas iniciativas convergen con la propuesta de Yoo (2025) de incluir secciones obligatorias de "AI Disclosure" en manuscritos y con la sugerencia de Teixeira da Silva (2024) de adoptar listas de verificación universales equivalentes a PRISMA o CONSORT. La revisión revela la emergencia de estándares interdisciplinares de transparencia, aunque aún fragmentados por área y editorial. Paralelamente, estudios como los de Hu et al. (2025) y Calderón y Herrera (2025) advierten sobre riesgos metodológicos derivados de la inestabilidad de respuestas ante la repetición de prompts y sobre la necesidad de reproducibilidad mediante flujos de trabajo con control de versiones y auditoría de cambios.

Los hallazgos muestran que la integración responsable de la IA en el ciclo de la investigación avanza hacia un modelo de rendición de cuentas compartida. La tendencia más consolidada es la de exigir divulgación detallada y trazabilidad, seguida de la institucionalización de protocolos de revisión ética y de publicación transparente. Persisten, sin embargo, brechas notables: ausencia de métricas comparables de impacto, desigualdad de recursos entre regiones y disciplinas, y falta de interoperabilidad entre sistemas de reporte. La evidencia disponible sugiere que el desarrollo de una arquitectura mínima de adopción responsable, basada en divulgación verificable, responsabilidad humana, control editorial y trazabilidad reproducible, constituye la vía más viable para equilibrar innovación y confianza en la comunicación científica.

Discusión

Esta revisión de alcance evidencia una convergencia sustantiva en torno a cómo integrar la IA generativa en el ciclo de la investigación sin erosionar la integridad científica. Los estudios sintetizados, aunque heterogéneos en disciplinas y métodos, coinciden en un principio rector: la IA aporta eficiencia y claridad en tareas de apoyo, sobre todo en redacción, edición lingüística y organización de contenidos, pero su valor depende de una arquitectura de gobernanza que asegure divulgación granular, supervisión humana indelegable, confidencialidad en procesos sensibles y trazabilidad reproducible. La literatura reciente ya no discute si emplear IA, sino bajo qué condiciones, en qué etapas, y con qué garantías de transparencia y responsabilidad.

En primer lugar, se consolida una noción de gobernanza práctica basada en la revelación detallada del uso de IA. Se amplía el estándar de las meras "declaraciones de uso" hacia esquemas más finos que identifican herramienta, versión, fecha de uso, alcance de la tarea (por ejemplo, revisión de estilo frente a generación de texto original), parámetros relevantes y verificación humana. Este desplazamiento responde a dos tensiones: la no estacionariedad de los modelos, que obliga a datar y especificar versiones, y la necesidad de reproducibilidad, que requiere documentar prompts, plantillas o procedimientos