Transparencia e integridad en el uso de la inteligencia artificial generativa en manuscritos científicos: la iniciativa GAMER

Transparency and integrity in the use of generative artificial intelligence in scientific manuscripts: the GAMER initiative

Antoni Sisó-Almirall[a]

Director de Investigación del Consorci d'Atenció Primària de Salut Barcelona-Esquerre (CAPSBE). Profesor titular de la Facultad de Medicina. Universidad de Barcelona. Coordinador del Máster IA en Investigación y Docencia en Ciencias de la Salud. Universidad de Barcelona. Barcelona (España). Presidente de la CAMFiC

CORREO ELECTRÓNICO:
Antoni Sisó-Almirall. asiso@clinic.cat

La inteligencia artificial generativa (IAG) ya forma parte del ciclo de vida de la investigación: ayuda a generar, estructurar y organizar textos; a crear imágenes y vídeos; a resumir evidencia; programar e incluso a generar nuevas hipótesis. Pero esta realidad también exige prudencia y unas normas de uso responsable. El *GAMER Statement* propone una lista de comprobación de nueve ítems para declarar y auditar el uso de herramientas generativas en investigación médica[1]. En Atención Primaria (AP) este estándar es especialmente necesario: trabajamos con datos muy sensibles, longitudinales y biográficos, tomando decisiones compartidas bajo incertidumbres, y cualquier opacidad metodológica erosiona la confianza. La IAG se caracteriza por generar texto nuevo, pero los modelos predicen en sus respuestas aquello que es probable, y no necesariamente aquello verdadero. A pesar de ofrecer resultados convincentes, estos pueden ser incorrectos, inventar referencias o sostener resultados engañosos fundamentados en bases de datos sesgadas, como ya se ha descrito recientemente[2].

Las recomendaciones sobre el uso de IAG en las publicaciones es dispar: algunas se limitan a pedir una declaración general, rechazando autorías que se atribuyan a la IA; otras exigen una declaración con detalles técnicos trazables[3], reinando la heterogeneidad editorial en las secciones de instrucciones a autores[4]. GAMER no prohíbe la IAG, pero exige transparencia sistemática y verificable, en la línea de propuestas recientes para un uso responsable de herramientas generativas en investigación[5].

GAMER se desarrolló mediante dos rondas Delphi: 51 expertos de 26 países con perfiles clínicos, metodológicos, éticos, tecnológicos y editoriales. El resultado es aplicable a cualquier tipo de estudio y cualquier fase del proyecto (diseño, recolección de datos, análisis y redacción). Se centra en herramientas genuinamente generativas (modelos de lenguaje, visuales o multimodales) y no regula otros tipos de herramientas no generativas como buscadores o traductores.

Las recomendaciones incluyen un *checklist* con definiciones y explicaciones para facilitar su aplicación. Y este enfoque se agradece, porque las guías de comprobación deben mejorar su reproducibilidad y rendición de cuentas, pero no a costa de añadir burocracia[6].

Los nueve ítems GAMER

GAMER se concreta en nueve requisitos completamente aplicables a cualquier ámbito de investigación (**tabla 1**), como en AP:

1. Declaración general de uso: indicar si se usó o no IAG en cualquier paso del estudio o manuscrito.
2. Especificación de herramienta y período: nombre, versión o fecha de lanzamiento, período exacto de uso y, cuando sea posible, si se utilizó interfaz y/o parámetros relevantes.
3. *Prompting* y trazabilidad: describir técnicas de *prompting* y, cuando sea posible, aportar los *prompts* y respuestas no editadas como material suplementario.
4. Modelos nuevos o ajustados: si se entrenó o afinó un modelo generativo, declarar el modelo original y su versión.
5. Rol de la IAG en todas las fases: detallar para qué se usó (ideación, protocolo, análisis, generación de código, redacción, tablas o figuras).
6. Localización en el manuscrito: señalar secciones o párrafos con contribución de IAG.
7. Verificación del contenido: explicar cómo se verificó y, si fue necesario, se corrigió lo generado; siempre recordando que la responsabilidad última sigue siendo humana[3].

El contenido de la Revista Clínica de Medicina de Familia está sujeto a las condiciones de la licencia de Creative Commons Reconocimiento-NoComercial-SinObraDerivada 4.0

Tabla 1. GAMER en 60 segundos: qué exige y cómo aplicarlo en Atención Primaria

ÍTEM GAMER	QUÉ EXIGE (MÍNIMO)	TRADUCCIÓN A LA PRÁCTICA
1. Uso de IAG	Sí/no	Declarar también «no uso»
2. Herramienta	Nombre + versión + fechas	Registrar versión y período
3. *Prompts*	Técnica + diálogos	Guardar *prompts*/respuestas y anexar
4. Modelo propio	Modelo base	Narrar el entrenamiento/afinamiento local
5. Rol	Para qué se usó	Diferenciar estilo vs. análisis/decisiones
6. Dónde	Secciones/párrafos	Señalar discusión, métodos, tablas, etc.
7. Verificación	Cómo se comprobó	*Fact-checking* + revisión de referencias
8. Privacidad	Salvaguardas	Desidentificar y controlar plataforma
9. Impacto	Influencia en conclusiones	Declarar si cambió la interpretación

8. Privacidad y confidencialidad: describir salvaguardas para proteger datos durante el uso de IAG; en nuestro ámbito exige procesos de anonimización, seudononimización, y control de plataformas[7].

9. Impacto en conclusiones: declarar si el uso de IAG influyó en la interpretación o en las conclusiones.

La aportación de GAMER en AP nos obliga a diferenciar usos de bajo riesgo (p. ej., corrección de estilo) de usos con capacidad de alterar resultados (p. ej., generación de código, extracción automatizada de datos, o síntesis analítica). y añade trazabilidad cuando la IAG participa en la construcción del manuscrito. En términos operativos, GAMER invita a planificar el uso de IAG antes de escribir: decidir qué tareas se permitirán (y cuáles no), quién revisará las salidas, dónde se archivarán *prompts* y respuestas, y cómo se protegerán los datos. En estudios con historias clínicas o entrevistas, el *checklist* es útil como recordatorio de «líneas rojas»: nunca introducir información identificable y documentar las salvaguardas.

Implicaciones prácticas para autores, revisores y editores

Para los investigadores, GAMER introduce hábitos de buenas prácticas: registrar versiones, conservar *prompts*, archivar respuestas, documentar verificaciones y definir qué datos pueden entrar (y en qué entorno). En proyectos con historias clínicas electrónicas, esto debe traducirse en declarar los procedimientos de anonimización o seudononimización con herramientas aprobadas institucionalmente y con la necesaria aprobación de los comités de ética, los cuales (dicho sea de paso), deberán ser competentes en saber evaluar el uso de modelos de IAG. Para revisores y editores, el *checklist* contribuye a despejar la duda («¿Se usó IA?») en preguntas verificables («¿Qué modelo?, ¿qué *prompts*?, ¿cómo se comprobó?») aplicando criterios consistentes: solicitar el *checklist* cumplimentado y exigir una trazabilidad proporcional al riesgo.

GAMER no sustituye guías específicas ya existentes tales como TRIPOD+AI[8] (que declara el uso transparente de modelos de predicción clínica que usan aprendizaje automático/IA), CONSORT-AI[9] (para ensayos clínicos que usan IA) o DECIDE-AI (sistemas de apoyo a la decisión con IA)[10]. Lo que hace GAMER es complementarlas, ofreciéndose como un denominador común amigable cuando usemos IAG en cualquier etapa de la investigación. No obstante, el nivel de exigencia irá en aumento, como lo demuestra la aparición de guías centradas en grandes modelos de lenguaje como TRIPOD-LLM[11].

En AP ya convivimos con *chatbots*. Comparativas entre actuaciones de médicos frente a *chatbots* a preguntas o situaciones clínicas están demostrando potencialidades que nos obligan a discutir sobre seguridad, sesgos y adecuación clínica y contextual[12,13]. En contraposición a alguna proclama política, la IAG no sustituirá a los profesionales: y mucho menos a las médicas y los médicos de familia, en cuyo día a día la creatividad y la interacción son difícilmente sustituibles. Sin embargo, los profesionales que utilicen correctamente la IAG sí podrían reemplazar en el futuro a aquellos que no la usen. Y en el ámbito de la publicación científica ocurre algo similar: la IAG puede elevar la calidad y la eficiencia, pero solo si se hace de forma visible, auditable y responsable. En cualquier ámbito de investigación y publicación, la inclusión de una lista de comprobación como GAMER nos ofrecerá un estándar mínimo de buenas prácticas y transparencia: una decisión de integridad personal e institucional para quienes la adopten.

BIBLIOGRAFÍA

1. Luo X, Tham YC, Giuffrè M, Ranisch R, Daher M, Lam K, et al. Reporting guideline for the use of Generative Artificial intelligence tools in Medical Research: the GAMER Statement. BMJ Evid Based Med. 2025;30:390-400.
2. Naddaf M. ChatGPT generates fake data set to support scientific hypothesis. Nature New Biol. 2023;623:895-896.
3. Flanagin A, Pirracchio R, Khera R, Berkwits M, Hswen Y, Bibbins-Domingo K. Reporting Use of AI in Research and Scholarly Publication—JAMA Network Guidance. JAMA. 2024;331:1096-8.
4. Ganjavi C, Eppler MB, Pekcan A, Biedermann B, Abreu A, Collins GS, et al. Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: bibliometric analysis. BMJ. 2024; 384:e077192.
5. Islam N, Van der Schaar M. Use of generative artificial intelligence in medical research. BMJ. 2024; 384:q119.

6. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med. 2010; 7:e1000217.

7. Wu X, Duan R, Ni J. Unveiling security, privacy, and ethical concerns of ChatGPT. Journal of Information and Intelligence. 2024; 2:102-15.

8. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024; 385:e078378.

9. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK, SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med. 2020; 26:1364-74.

10. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med. 2022; 28:924-33.

11. Gallifant J, Afshar M, Ameen S, Aphinyanaphongs Y, Chen S, Cacciamani G, et al. The TRIPOD-LLM reporting guideline for studies using large language models. Nat Med. 2025; 31:60-69.

12. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. JAMA Intern Med. 2023; 183:589-96.

13. Hayat H, Kudrautsau M, Makarov E, Melnichenko V, Tsykunou T, Varaksin P, et al. Toward the Autonomous AI Doctor: Quantitative Benchmarking of an Autonomous Agentic AI Versus Board-Certified Clinicians in a Real World Setting. medRxiv (preprint) 2025.07.14.25331406. doi: 10.1101/2025.07.14.25331406.