

Ciencia, bienes comunes y soberanía digital: Hacia una inteligencia artificial gobernada por la comunidad

Science, common goods and digital sovereignty: Towards a community-governed artificial intelligence

Humberto Debat

Citación recomendada:

Debat, Humberto (2026). "Ciencia, bienes comunes y soberanía digital: Hacia una inteligencia artificial gobernada por la comunidad [Science, common goods and digital sovereignty: Towards a community-governed artificial intelligence]". *Infonomy*, 4(3) e26018.

<https://doi.org/10.3145/infonomy.26.018>

Artículo recibido: 26-03-2026

Artículo aprobado: 27-05-2026



Humberto Debat

<https://orcid.org/0000-0003-3056-3739>

<https://directorioexit.info/ficha7085>

Instituto Nacional de Tecnología Agropecuaria

Argentina

University of Toronto

Canadá

debat.humberto@inta.gob.ar

Agradecimiento

Este artículo nace a partir de un comentario de la siempre brillante Dominique Babini.

Resumen

La rápida integración de grandes modelos de lenguaje (LLM, por sus siglas en inglés) e inteligencia artificial generativa en los flujos de trabajo científicos ha introducido una asimetría estructural: los instrumentos que median de forma creciente la síntesis del conocimiento son propiedad de entidades comerciales, que los gobiernan y optimizan según estructuras de incentivos divergentes respecto de las normas epistémicas de la ciencia. En este artículo



sostengo que los marcos regulatorios basados en principios éticos, si bien son necesarios, resultan insuficientes para salvaguardar la integridad del registro científico o para garantizar la participación equitativa de los investigadores en países de ingresos bajos y medios. A través de la síntesis de evidencia proveniente de la utilización de inteligencia artificial (IA) en investigación clínica, análisis de ciencia abierta, e investigación en salud pública, este artículo pone de relieve los riesgos poblacionales asociados a un conocimiento generado o mediado por intereses comerciales. Para mitigar dichos riesgos, argumento que los organismos científicos internacionales deben liderar la transición hacia una infraestructura de IA soberana y de propiedad comunitaria. Sitúo esta propuesta en el contexto del impulso institucional generado por el *Pacto Digital Global de las Naciones Unidas*, el establecimiento del *Panel Científico Internacional Independiente sobre Inteligencia Artificial* y el incipiente *Diálogo Global sobre Gobernanza de la IA*. Apoyándome en los precedentes del *CERN* y la *Fundación Wikimedia*, sostengo que la comunidad investigadora global debe traducir la voluntad política en soberanía computacional concreta. Diseñada como un bien común abierto (*Open Commons*), dicha infraestructura trascendería la mera utilidad en investigación para constituirse en la fuerza impulsora de un ecosistema de conocimiento global gobernado por la comunidad.

Palabras clave

Ciencia; Conocimiento global; Bien común; Soberanía digital; Inteligencia artificial; Gobernanza; Comunidad; Grandes modelos de lenguaje; Marcos regulatorios; Integridad; Equidad; Organismos científicos internacionales; *CERN*; *Unesco*; *ONU*; *Naciones Unidas*; Infraestructura; *Fundación Wikimedia*.

Abstract

The rapid integration of large language models (LLMs) and generative artificial intelligence into scientific workflows has introduced a structural asymmetry: the instruments that increasingly mediate knowledge synthesis are owned by commercial entities, which govern and optimize them according to incentive structures that diverge from the epistemic norms of science. In this paper, I argue that regulatory frameworks grounded in ethical principles, while necessary, are insufficient to safeguard the integrity of the scientific record or to ensure the equitable participation of researchers in low- and middle-income countries. Through a synthesis of evidence drawn from the use of artificial intelligence (AI) in clinical research, open science analysis, and public health research, this paper highlights the population-level risks associated with knowledge generated or mediated by commercial interests. To mitigate these risks, I argue that international scientific bodies must lead the transition toward sovereign, community-owned AI infrastructure. I situate this proposal within the institutional momentum generated by the *United Nations Global Digital Compact*, the establishment of the *Independent International Scientific Panel on Artificial Intelligence*, and the nascent *Global Dialogue on AI Governance*. Drawing on the precedents set by *CERN* and the *Wikimedia Foundation*, I contend that the global research community must translate political will into concrete computational sovereignty. Designed as an Open Commons, such infrastructure would transcend mere research utility to become the driving force of a community-governed global knowledge ecosystem.

Keywords

Science; Global knowledge; Common good; Digital sovereignty; Artificial intelligence; AI; Governance; Community; Big language models; Regulatory frameworks; Integrity; Equity; International scientific organizations; *CERN*; *UNESCO*; *UN*; *United Nations*; Infrastructure; *Wikimedia Foundation*.

1. Introducción

La integración de los grandes modelos de lenguaje (LLM) y la inteligencia artificial generativa en la práctica científica marca una transición de magnitud comparable a la digitalización de la comunicación académica y al surgimiento de las bases de datos de investigación en línea. Análisis recientes sobre los denominados «modelos fundacionales» (*foundation models*) subrayan que estos sistemas constituyen un cambio tecnológico general con implicancias de largo plazo en todos los dominios científicos, que altera de manera fundamental el modo en que se genera, se procesa y se aplica el conocimiento (**Bommasani et al.**, 2021). Sin embargo, a diferencia de aquellas transformaciones previas, el cambio actual se define por un desequilibrio estructural: los instrumentos que median de forma creciente el descubrimiento, la síntesis y la interpretación del conocimiento son desarrollados, poseídos y gobernados por corporaciones privadas, y no por la propia comunidad científica. Como resultado, la aspiración de construir un sistema de producción de conocimiento abierto, equitativo y globalmente inclusivo se despliega sobre un sustrato tecnológico que es, en su esencia, propietario.

Este desequilibrio no es incidental, sino sistémico. Al tiempo que los LLM se integran con rapidez en dominios como el razonamiento clínico, la educación y la interpretación de datos, un creciente cuerpo de evidencia pone de manifiesto limitaciones fundamentales en su fiabilidad. Estudios empíricos demuestran que las respuestas de los LLM pueden variar incluso en condiciones idénticas, lo que suscita preocupaciones sobre la repetibilidad y la reproducibilidad en contextos científicos y médicos (**Shyr et al.**, 2025). Evaluaciones sistemáticas muestran, además, que estos sistemas siguen siendo vulnerables a las alucinaciones, incluida la generación con aparente certeza de información fabricada o engañosa, aun en aplicaciones de alto riesgo como el apoyo a la toma de decisiones clínicas y la redacción científica (**Omar et al.**, 2025; **Chelli et al.**, 2024; **Sun et al.**, 2025). Los análisis cuantitativos de la síntesis bibliográfica asistida por LLM revelan altas tasas de alucinación y escasa precisión referencial, lo que subraya los riesgos de emplear estos sistemas como herramientas primarias de generación de conocimiento (**Chelli et al.**, 2024). Trabajos complementarios en seguridad de la IA clínica demuestran asimismo que incluso modelos en apariencia eficaces pueden exhibir discrepancias de fidelidad entre los resultados generados y los datos de referencia, con implicancias directas para la seguridad del paciente y la validez científica

Los instrumentos que median de forma creciente el descubrimiento, la síntesis y la interpretación del conocimiento son desarrollados, poseídos y gobernados por corporaciones privadas, y no por la propia comunidad científica

(Asgari et al., 2025). De manera más general, la implementación de los LLM en contextos clínicos evidencia preocupaciones persistentes en cuanto a la fiabilidad, los estándares de evaluación y la integración segura en los flujos de trabajo (Shool et al., 2025). Los metaanálisis sobre la colaboración entre humanos y LLM sugieren que, si bien estos modelos pueden aportar mejoras en el rendimiento en determinados entornos, tales mejoras van acompañadas de una incertidumbre y variabilidad considerables en condiciones del mundo real (Wang et al., 2026). En conjunto, estos hallazgos indican que la integración de los LLM en los flujos de trabajo científicos ha trascendido la logística de su adopción para convertirse en un desafío fundamental a la estabilidad epistémica.

De forma concurrente, el epicentro institucional del desarrollo de la IA se ha desplazado de manera decisiva hacia la industria, concentrando la experiencia, los recursos computacionales y la capacidad de desarrollo de modelos fuera del ámbito académico. La consecuencia es históricamente inédita: las principales herramientas epistémicas de la ciencia ya no se producen en instituciones cuyo mandato central es la generación de conocimiento público. En su lugar, están moldeadas por estructuras de incentivos alineadas con la competencia de mercado, el despliegue de productos y la captación de usuarios. Esta divergencia introduce una tensión estructural entre las normas de la ciencia, transparencia, reproducibilidad y verificación acumulativa, y la lógica operativa de los sistemas comerciales de IA.

La respuesta regulatoria predominante frente a esta transformación se ha articulado en torno a un paradigma que prioriza la ética (*ethics-first*) (Vedder; Nisevic; Fleerackers, 2025). Si bien necesario, este enfoque resulta cada vez más insuficiente. Las directrices éticas y los mecanismos de supervisión *ex post* no pueden abordar plenamente un problema cuya naturaleza es infraestructural. Cuando los sistemas subyacentes permanecen opacos, mutables y gobernados externamente, la capacidad de la comunidad científica para garantizar la integridad de su propia producción de conocimiento se ve inherentemente afectada. La cuestión, por tanto, no se reduce a cómo debe utilizarse la IA en la ciencia, sino a quién controla los sistemas a través de los cuales se media cada vez más el conocimiento científico. Este desafío es particularmente acuciante en el contexto de las desigualdades globales en la producción de conocimiento. La evidencia proveniente de aplicaciones clínicas y biomédicas muestra que las limitaciones de los LLM, incluyendo la alucinación, el sesgo y la inconsistencia, pueden afectar de manera desproporcionada a aquellos contextos donde los recursos de validación y los datos específicos del dominio son escasos (Omar et al., 2025; Sun et al., 2025; Shool et al., 2025). En tales entornos, la dependencia de sistemas propietarios profundiza el riesgo de amplificar las disparidades preexistentes en el acceso epistémico y la participación.

La cuestión no se reduce a cómo debe utilizarse la IA en la ciencia, sino a quién controla los sistemas a través de los cuales se media cada vez más el conocimiento científico

Para abordar este desafío, propongo un desplazamiento desde la regulación aislada hacia la construcción institucional. En concreto, las organizaciones científicas internacionales deben coordinar el desarrollo de una infraestructura de IA soberana y de

propiedad comunitaria, diseñada explícitamente para el uso científico. Dicha infraestructura alinearía las capacidades tecnológicas con las normas epistémicas de la ciencia, garantizando al mismo tiempo un acceso y una participación equitativos entre regiones. Lejos de ser una proposición puramente teórica, este argumento emerge en un momento de acelerado compromiso multilateral con la gobernanza de la IA, que genera una oportunidad institucional concreta para que la comunidad científica ejerza agencia colectiva sobre el futuro de sus sistemas de conocimiento.

2. El desajuste entre la optimización comercial y la integridad científica

El argumento central a favor de un modelo de propiedad comunitaria descansa sobre la tensión entre la optimización comercial de propósito general y la precisión científica de dominio específico. Los LLM comerciales se entrenan mediante técnicas como el aprendizaje por refuerzo a partir de retroalimentación humana (*Reinforcement Learning from Human Feedback*, RLHF) para optimizar una «utilidad» genérica y el compromiso (*engagement*) del usuario en aplicaciones diversas. Si bien esto produce herramientas de propósito general útiles, tales objetivos de optimización pueden entrar en conflicto con los requisitos de precisión de la investigación científica. Para el usuario común, una alucinación plausible puede suponer un inconveniente informativo menor; para el investigador, constituye una fuente de contaminación epistémica que conduce a inferencias erróneas. Las evaluaciones de los LLM en aplicaciones clínicas documentan un control inconsistente del contenido alucinado, exceso de confianza en los resultados, sesgos heredados de los datos de entrenamiento y brechas de responsabilidad no resueltas (**Wang et al.**, 2026; **Mansoor et al.**, 2026). La opacidad de estos sistemas socava la transparencia esencial para la reproducibilidad científica (**Ma et al.**, 2025). En la predicción de resultados quirúrgicos, la integración clínica exitosa de la IA exige una validación externa robusta y el intercambio transparente de modelos entrenados para garantizar la reproducibilidad y la generalizabilidad entre cohortes diversas de pacientes (**Limon et al.**, 2025), requisitos que los sistemas propietarios no están estructuralmente destinados a satisfacer.

De forma crítica, el problema trasciende los fallos de modelos individuales. Un análisis exploratorio reciente sobre la intersección entre la IA generativa y la ciencia abierta, estructurado en torno a la taxonomía de la *Unesco*, concluyó que, si bien la IA generativa podría impulsar objetivos clave de la ciencia abierta, ampliando el acceso al conocimiento, permitiendo un uso eficiente de la infraestructura y enriqueciendo el diálogo entre sistemas de conocimiento, podría simultáneamente comprometer la integridad, la equidad, la reproducibilidad y la fiabilidad de la investigación (*Unesco*, 2025). Los autores enfatizan que resultan imprescindibles las verificaciones, validaciones y evaluaciones críticas al incorporar la IA generativa en los flujos de trabajo científico. Tales verificaciones son, por definición, imposibles cuando el modelo es una caja negra comercial. El panorama regulatorio no ha permanecido inactivo. La naturaleza de caja negra de los modelos de IA representa una barrera significativa para la aceptación regulatoria en múltiples dominios, lo que refuerza el argumento a favor de sistemas abiertos, explicables y auditables (**Hosseini et al.**, 2025). No obstante, la regulación avanza al ritmo de la deliberación de comités; las actualizaciones de los modelos, al ritmo de los informes trimestrales de resultados. Cuando los investigadores no pueden especificar qué difiere entre versiones de un modelo que emplearon, o cuando el

comportamiento de un modelo cambia entre la consulta y la revisión por pares, los fundamentos de la verificación se desmoronan.

La crisis de reproducibilidad en la ciencia es anterior a la IA, pero esta amenaza con agravarla a una escala sin precedentes. Estamos asistiendo a una aceleración en tiempo real del cierre de la investigación corporativa en IA: *DeepMind* ha impuesto embargos de publicación de seis meses para proteger su ventaja competitiva; *OpenAI* ha transitado de la publicación con revisión por pares a los anuncios de productos propietarios; y *Meta* ha desmantelado su unidad de *Investigación Fundamental en IA* justo cuando esta se convierte en una industria billonaria (**Etchemendy et al.**, 2025). Esto constituye una falla de mercado en la formación de talento en IA y la producción de conocimiento público, y evidencia la necesidad de que las universidades reafirmen su papel mediante centros de investigación distribuidos y transcontinentales que compartan liderazgo, datos, capacidad computacional y talento profesional (**Etchemendy et al.**, 2025). La solución no puede recaer exclusivamente en las universidades del Norte Global; requiere el tipo de infraestructura gobernada internacionalmente y de propiedad comunitaria que garantice la participación equitativa entre regiones.

3. La carga amplificada sobre el Sur Global

El problema del desajuste es más agudo para los investigadores en países de ingresos bajos y medios (PIEMEDS). Los modelos comerciales se entrenan predominantemente con conjuntos de datos que reflejan el poder económico de sus mercados objetivo. De los aproximadamente 7.000 idiomas hablados en el mundo, solo unos 20 disponen de recursos digitales suficientes para un procesamiento robusto del lenguaje natural, lo que significa que más del 99% de las lenguas del mundo son tratadas como casos marginales de «bajos recursos» (**Hedderich et al.**, 2021). Cuando los investigadores en PIEMEDS consultan modelos comerciales sobre contextos locales, ya se trate de resiliencia agrícola, epidemiología regional o sistemas de conocimiento tradicionales, pueden recibir respuestas derivadas de aproximaciones basadas en datos occidentales, o encontrar negativas por insuficiencia del contexto de entrenamiento. Esta limitación técnica es consecuencia estructural directa de una lógica de mercado que prioriza las regiones de altos ingresos, dejando escaso incentivo comercial para curar datos de entrenamiento de alta calidad destinados a poblaciones marginadas lingüística o geográficamente. Los desafíos van más allá del idioma. Los obstáculos fundamentales para la adopción de la IA en la atención sanitaria de los PIEMEDS incluyen la conectividad a internet deficiente, la carencia de equipamiento, la escasa motivación del personal y la dirección, la distribución desigual de recursos y las preocupaciones éticas en torno al intercambio de datos (**Kaushik et al.**, 2025). Los comités de ética de investigación en los PIEMEDS enfrentan el desafío adicional de evaluar herramientas de IA desarrolladas sin considerar los diferenciales históricos de poder en la investigación colaborativa internacional, consideraciones que no se delegan fácilmente a sistemas automatizados (**Moodley; Malpani; Reis**, 2025).

Resulta significativo que las prácticas de datos abiertos hayan demostrado efectos mensurables sobre la equidad. La participación de autores procedentes de PIEMEDS es sustancialmente mayor entre quienes utilizan conjuntos de datos abiertos en comparación con datos propietarios, incluidas las posiciones protagónicas de primer y último autor (**Charpignon; Celi; Cobanaj**, 2024). Esto sugiere que la infraestructura

abierta puede apoyar directamente la participación de regiones subrepresentadas, y que la elección de la arquitectura infraestructural no es neutra respecto a la distribución de la agencia epistémica. Solo una infraestructura pública, regida por un mandato de equidad epistémica en lugar de lucro, puede justificar la inversión sostenida requerida para digitalizar y respetar los sistemas de conocimiento desatendidos. La alternativa, la dependencia perpetua de sistemas comerciales cuyas prioridades de entrenamiento están determinadas por el potencial de ingresos, consolida lo que equivale a una nueva forma de colonialismo epistémico, mediada ya no solo por las revistas y las barreras lingüísticas, sino por los propios instrumentos de síntesis del conocimiento. Esta dinámica ha sido teorizada como «tecnocolonialidad», una condición en la que el uso de la tecnología reproduce patrones coloniales de poder, control y dominación del conocimiento.

Como ejemplo, Mboa-Nkoudou identifica cuatro dimensiones a través de las cuales la IA perpetúa las estructuras coloniales en África (**Mboa-Nkoudou, 2023**):

- el discurso tecno-utópico que enmascara fracasos tras promesas inspiradas en realidades foráneas;
- la colonialidad del conocimiento sostenida mediante conjuntos de datos de entrenamiento centrados en occidente;
- la transferencia acrítica de tecnología que impone sistemas estructuralmente ajenos sin contextualización; y
- el extractivismo neocapitalista de datos, en el que se explotan datos de las naciones en desarrollo sin compensación adecuada.

Mboa-Nkoudou no aboga por el rechazo de la IA, sino por su «apropiación descolonizada»: la capacidad de impugnar herramientas que no se corresponden con las realidades locales y, cuando sea posible, de adaptar, reconvertir y recrear los sistemas foráneos para atender las necesidades locales. Esta es precisamente la lógica que sustenta el modelo de infraestructura de propiedad comunitaria que aquí se propone: no la renuncia a la IA, sino la participación soberana en su gobernanza y desarrollo.

4. La paradoja fiduciaria: pagar por nuestra propia expropiación

El modelo económico vigente genera una profunda paradoja fiduciaria para los financiadores públicos de la investigación. Los consejos de investigación y las agencias de financiación destinan miles de millones anuales a apoyar la investigación, una parte considerable de la cual se publica bajo mandatos de acceso abierto para garantizar su condición de bien público. Este corpus financiado con fondos públicos es incorporado, a menudo sin permiso explícito ni compensación, a los conjuntos de datos de entrenamiento por parte de empresas tecnológicas privadas. Estas mismas empresas venden posteriormente a las universidades e institutos públicos de investigación la inteligencia sintetizada a partir de dichos datos, a través de pagos de API y suscripciones empresariales. El ciclo transfiere valor efectivo del sector público para subsidiar la extracción de rentas (*rent-seeking*) del sector privado. Esta dinámica presenta paralelismos estructurales con la crisis de las suscripciones a revistas científicas que catalizó el movimiento de acceso abierto, pero con consecuencias de un orden de magnitud superior: mientras las editoriales controlaban la distribución del conocimiento, las empresas de IA se están posicionando para controlar su síntesis.

La transición desde el reconocimiento de esta paradoja estructural hasta su resolución depende de una reevaluación crítica de los umbrales fiscales y técnicos reales de la infraestructura soberana. Sin embargo, desarrollos recientes sugieren que los requisitos de capital para la IA de frontera pueden ser inferiores a lo previamente supuesto. A comienzos de 2025, diversas innovaciones en eficiencia demostraron que modelos de alta capacidad podían entrenarse con recursos computacionales significativamente menores en comparación con las generaciones anteriores. Si bien las estimaciones iniciales de costes de entrenamiento resultaron sustancialmente subestimadas al incorporar la infraestructura completa y las inversiones previas en investigación, con costes reales probablemente en el orden de las decenas de millones de dólares, la trayectoria indica, no obstante, que el sector público enfrenta una cuestión de asignación de recursos más que de imposibilidad fiscal (*United Nations General Assembly*, 2024). Mediante la coordinación de una reasignación estratégica de los gastos nacionales en investigación, actualmente dispersos entre licencias de software comercial, suscripciones a bases de datos propietarias y cargos por procesamiento de artículos (*Article Processing Charges*), las instituciones podrían movilizar el capital necesario para construir una infraestructura computacional pública y federada. El paralelismo con las inversiones nacionales en infraestructura soberana de IA resulta ilustrativo: la iniciativa *AI Factories* de la UE ha comprometido más de 4.200 millones de euros en el despliegue de capacidad computacional soberana solo durante 2024-2025, y Canadá ha lanzado un *Programa de Infraestructura de Computación de IA Soberana* orientado explícitamente a la infraestructura pública como bien común. En lugar de crear un modelo institucional enteramente nuevo, la comunidad científica puede reconvertir los marcos soberanos que los gobiernos nacionales ya están estableciendo para la misión transnacional más amplia, la de la ciencia.

El sector público debe poseer los medios de síntesis del conocimiento, no meramente los datos brutos. Un ejemplo ilustrativo de este enfoque, ya en marcha, es *Latam-GPT*, una iniciativa colaborativa coordinada por el *Centro Nacional de Inteligencia Artificial (CENIA)* de Chile, en la que participan cerca de 200 profesionales de más de 65 instituciones en 15 países. Puesto en marcha en febrero de 2026, *Latam-GPT* está construyendo un gran modelo de lenguaje abierto entrenado con un corpus de más de 300.000 millones de *tokens* extraídos de fuentes regionales: sentencias judiciales, registros de bibliotecas, libros de texto escolares y otros conjuntos de datos que reflejan las realidades lingüísticas, culturales e institucionales de América Latina y el Caribe (*Latam-GPT*, s. f.; **Levy-Yeyati**, 2025). El proyecto se concibe explícitamente como un bien público tecnológico y no como un producto comercial, y sus tres pilares declarados, el desarrollo de capacidades técnicas locales, la corrección de la brecha de representación regional en los datos globales de entrenamiento de IA (el español constituye aproximadamente el 4% y el portugués entre el 2% y el 3% de los corpus de entrenamiento típicos), y la soberanía tecnológica, se corresponden directamente con los argumentos desarrollados en este artículo. Cabe destacar que la lista de participantes de *Latam-GPT* incluye no solo universidades y centros de investigación, sino también ministerios gubernamentales, bancos multilaterales de desarrollo (el *BID*, la *CAF* y la *OEA*) y *Wikimedia Chile*, lo que ilustra el tipo de colaboración intersectorial y gobernada por la comunidad que requeriría una *Red Global de Computación Científica (Global Science Compute Grid)*. El proyecto demuestra que la infraestructura soberana

de IA no necesita comenzar en la frontera tecnológica; puede comenzar con la construcción deliberada de sistemas de conocimiento representativos y gestionados localmente que cubran las carencias que los modelos comerciales carecen de incentivos para atender.

5. El momento multilateral: del Pacto Digital Global al Panel Científico sobre IA

En septiembre de 2024, la *Asamblea General* de las *Naciones Unidas* adoptó el *Pacto para el Futuro* (A/RES/79/1), que incluyó como primer anexo el *Pacto Digital Global* (*Global Digital Compact*), un marco integral para la cooperación digital internacional (*United Nations General Assembly, 2024*). El *Pacto* compromete a los Estados miembros a «mejorar la gobernanza internacional de la inteligencia artificial en beneficio de la humanidad» y solicita específicamente el desarrollo de «software de código abierto, datos abiertos, modelos de inteligencia artificial abiertos, estándares abiertos y contenido abierto» como bienes públicos digitales. El *Pacto* establece como meta para 2030 la diseminación de modelos y estándares de IA de código abierto diseñados para beneficiar a la sociedad en su conjunto (*United Nations General Assembly, 2024*). Al hacerlo, adopta la Recomendación de la *Unesco* sobre la Ética de la Inteligencia Artificial como marco ético subyacente para estos bienes comunes digitales (*Unesco, 2021*). No se trata de declaraciones meramente aspiracionales. Son compromisos adoptados por consenso al nivel de Jefes de Estado y de Gobierno, y establecen un marco normativo explícito dentro del cual la propuesta de una infraestructura científica de IA de propiedad comunitaria no solo es compatible, sino esperada. La arquitectura institucional para operacionalizar estos compromisos ya está tomando forma. El 26 de agosto de 2025, la *Asamblea General* adoptó la resolución A/RES/79/325, que establece dos mecanismos: el *Panel Científico Internacional Independiente sobre Inteligencia Artificial* y el *Diálogo Global sobre Gobernanza de la IA* (*United Nations General Assembly, 2025*). El *Panel Científico*, integrado por 40 expertos seleccionados entre más de 2.600 candidaturas procedentes de más de 140 países, designados por la *Asamblea General* en febrero de 2026, tiene el mandato de elaborar evaluaciones científicas anuales basadas en evidencia sobre las oportunidades, los riesgos y los impactos de la IA (*United Nations General Assembly, 2025; Independent International Scientific Panel on AI, s. f.*). Los miembros desempeñan sus funciones a título personal, con salvaguardas frente a conflictos de intereses que garantizan su independencia respecto de gobiernos, corporaciones y organizaciones internacionales por igual.

El mandato del *Panel* es amplio: abarca desde los sistemas de frontera hasta los impactos sociales, e incluye el desarrollo de IA segura y fiable, los modelos y datos de código abierto, la transparencia y la supervisión humana, y la protección de los derechos humanos en el ámbito de la IA (*Independent International Scientific Panel on AI, s. f.*). Como enfatizó el Secretario General de la *ONU*, António Guterres, en la reunión inaugural del *Panel*, el mundo necesita urgentemente

«una comprensión global y compartida de la inteligencia artificial; basada no en la ideología, sino en la ciencia; no en noticias falsas, sino en el conocimiento» (**Guterres, 2026**).

Tres características específicas de este organismo revisten particular relevancia para la propuesta de una red global de computación científica para IA. En primer lugar, el

mandato del *Panel* es explícitamente analítico y no prescriptivo en materia de políticas, centrado en proporcionar evaluaciones rigurosas y basadas en evidencia en lugar de imponer estándares. Esta orientación crea un punto de entrada crucial para que la comunidad científica presente propuestas técnicas de infraestructura soberana ante un organismo capacitado para evaluarlas primordialmente en función de sus méritos científicos. Las conclusiones de tales evaluaciones pueden ser luego elevadas de forma sistemática al *Diálogo Global*, donde los Estados miembros y las partes interesadas internacionales se congregan para traducir la evidencia en gobernanza.

En segundo lugar, el *Panel* ha sido diseñado para funcionar como un «sistema de alerta temprana y motor de evidencia» (*Independent International Scientific Panel on AI*, s. f.). Una infraestructura de IA gobernada por la comunidad, caracterizada por el versionado transparente de modelos, resultados reproducibles y una cobertura multilingüe equitativa, representa precisamente el tipo de innovación institucional que un panel basado en la evidencia está en condiciones de validar. Si la argumentación científica se articula con el rigor suficiente, el *Panel* deviene el mecanismo principal para demostrar cómo dichas arquitecturas de interés público mitigan los riesgos de la opacidad propietaria. Finalmente, el compromiso del *Panel* con la inclusividad constituye un mandato estructural más que un gesto retórico. Con una representación equilibrada en los cinco grupos regionales de la *ONU* y la inclusión explícita de expertos tanto de países en desarrollo como de países desarrollados, la composición del *Panel* garantiza que las preocupaciones de los investigadores en PIEMEDS en materia de equidad epistémica y soberanía de datos permanezcan en el centro de sus deliberaciones. Este modelo de gobernanza diverso proporciona el entorno institucional idóneo para abogar por una infraestructura que sirva a los intereses colectivos de la comunidad investigadora global. La comunidad científica debe reconocer este momento multilateral por lo que verdaderamente es: no un sustituto de la acción comunitaria, sino un entorno propicio que podría transformar lo que hasta ahora ha sido una propuesta marginal en una prioridad institucional de primer orden. No obstante, la ventana de oportunidad es finita. El informe inaugural del *Panel* está programado para su presentación en el primer *Diálogo Global sobre Gobernanza de la IA* en julio de 2026. Si la comunidad científica aspira a influir en el encuadre de este documento fundacional, las especificaciones técnicas y los marcos de gobernanza para una IA científica de propiedad comunitaria deben articularse ahora.

6. La gobernanza como antídoto frente al borrado epistémico

El *Consejo Internacional de la Ciencia (ISC)*, por sus siglas en inglés) y la *Unesco* han identificado correctamente la «Ciencia Abierta» como un pilar del desarrollo futuro. La *Unesco* ha sido el organismo multilateral más proactivo en el establecimiento de fundamentos normativos para la gobernanza de la IA. Su *Recomendación sobre la Ética de la Inteligencia Artificial* fue el primer instrumento normativo global en materia de ética de la IA, al establecer principios de transparencia, equidad y supervisión humana junto con áreas de acción política que abarcan la gobernanza de datos, la educación, la investigación, la salud y el medio ambiente (*Unesco*, 2021). La *Recomendación* insta explícitamente a los Estados miembros a promover repositorios abiertos para datos y código fuente financiados con fondos públicos, y a invertir en la creación de conjuntos de datos abiertos y confiables, disposiciones que respaldan directamente el modelo de infraestructura aquí propuesto. Más recientemente, la *Unesco* ha

reconocido que la opacidad de los sistemas de IA y el predominio de las empresas privadas en el avance de esta tecnología plantean obstáculos significativos, y que la integración de la IA debe alinearse con los principios de la ciencia abierta para garantizar la replicabilidad, la credibilidad, la equidad y la confiabilidad (Unesco, 2025).

Sin embargo, la «Ciencia Abierta» en la era de la IA no puede limitarse a publicaciones y repositorios de datos abiertos; debe extenderse a los modelos de pesos abiertos (*open model weights*) y a la infraestructura computacional abierta. Si los algoritmos que seleccionan sintetizan y recuperan el conocimiento científico permanecen opacos y variables a discreción de la administración corporativa, la ciencia misma deviene irreproducible.

No es posible construir conocimiento acumulativo sobre instrumentos cuyos parámetros cambian sin previo aviso ni documentación. El *Pacto Digital Global* proporciona un respaldo normativo explícito a esta posición. Su compromiso con los bienes públicos digitales que se ajusten a la privacidad y otras leyes, estándares y mejores prácticas internacionales aplicables, y que no causen daño (*United Nations General Assembly, 2024*), abarca los modelos de IA abiertos como categoría de bien público. Su compromiso adicional de «promover la adopción de estándares abiertos e interoperabilidad para facilitar el uso de bienes públicos digitales en diferentes plataformas y sistemas» (*United Nations General Assembly, 2024*) apunta directamente a los requisitos de interoperabilidad de una infraestructura computacional científica federada. Un análisis reciente enmarca este desafío en términos estructurales: el desarrollo de un ecosistema abierto y una infraestructura integral para la IA en la ciencia podría incidir en el avance científico de manera tan profunda como la iniciativa de la red global de telecomunicaciones de alta velocidad de finales del siglo XX configuró el desarrollo de internet (Pei; Huang, 2025). Los autores argumentan que los recursos computacionales, los datos y los modelos deberían integrarse en un sistema unificado de provisión pública, que funcione «de manera similar a la electricidad o internet» como un bien público esencial. El proyecto piloto del *National AI Research Resource* (NAIRR) en EE. UU. (Siefert_McCanse, 2025) y el proyecto *Artificial Intelligence for the European Open Science Cloud* (AI4EOSC) de la UE (Trana, 2025) representan pasos iniciales en esta dirección, pero su escala sigue siendo limitada y geográficamente circunscrita al Norte Global.

Si los algoritmos que seleccionan sintetizan y recuperan el conocimiento científico permanecen opacos y variables a discreción de la administración corporativa, la ciencia misma deviene irreproducible

7. El modelo del CERN: un precedente para la infraestructura científica de propiedad comunitaria

Un LLM de propiedad comunitaria para la ciencia operaría bajo un marco de gobernanza análogo al del CERN o al de los protocolos fundacionales de internet. El CERN representa un modelo probado de colaboración internacional pacífica fundamentado en la transparencia, la apertura y la inclusión (Bona; Schopper; Spiro, 2016). Con aproximadamente 6.000 miembros distribuidos en 60 países, incluida una participación significativa del Sur Global, el éxito del CERN traslada el concepto de infraestructura a gran escala gobernada por la comunidad desde el ámbito de lo utópico al dominio de lo demostrablemente operativo. Recientemente, el CERN ha extendido su mandato de ciencia abierta de forma explícita a la infraestructura digital, estableciendo una

Open Science Office y alojando plataformas como el *Portal de Datos Abiertos del CERN*, *Zenodo* para resultados de investigación multidisciplinarios y el *Repositorio de Hardware Abierto (CERN, s. f.)*. Estos esfuerzos demuestran que la gobernanza comunitaria puede extenderse desde los aceleradores físicos hasta

los sistemas de conocimiento digital, una transición conceptualmente directa aunque organizativamente exigente. De manera decisiva, el modelo de gobernanza del *CERN* garantiza el versionado transparente y la toma de decisiones colectiva. Una IA científica de propiedad comunitaria permitiría de igual modo a los investigadores citar el estado exacto y trazable del modelo utilizado en su trabajo, posibilitando así la reproducibilidad. Podría ajustarse con «resultados negativos» para combatir el sesgo de publicación, optimizarse para detectar conexiones interdisciplinarias y entrenarse con corpus multilingües que reflejen la plena diversidad de la producción de conocimiento científico, inversiones que carecen de justificación comercial pero que cumplen funciones científicas esenciales. La analogía se asienta en una realidad material compartida: al igual que la física de partículas, la escala de la IA de frontera requiere instrumentos que exceden la capacidad de cualquier nación individual. La misma lógica se aplica a la infraestructura de IA para la ciencia: los recursos computacionales, los conjuntos de datos curados y las estructuras de gobernanza necesarios para construir y mantener un sistema de IA científica fiable superan la capacidad de las instituciones individuales e incluso de las naciones individuales, pero se hallan al alcance de un esfuerzo internacional coordinado.

Al igual que la física de partículas, la escala de la IA de frontera requiere instrumentos que exceden la capacidad de cualquier nación individual

Un segundo precedente relevante lo constituyen la *Fundación Wikimedia (Heracleous; Gößwein; Beaudette, 2018)* y su proyecto insignia, *Wikipedia (Voss, 2005)*. Durante casi veinticinco años, *Wikipedia* ha demostrado que una plataforma sin ánimo de lucro, gobernada por la comunidad, puede producir y mantener la enciclopedia más extensa de la historia de la humanidad, una que sirve como piedra angular de la infraestructura global del conocimiento y, no de manera incidental, como componente central de prácticamente todos los conjuntos de datos de entrenamiento de IA. La gobernanza de *Wikipedia* se encuentra estructuralmente alineada con los requisitos epistémicos de la ciencia, al priorizar la verificabilidad, la neutralidad y la transparencia a través de una comunidad global de editores que operan bajo políticas determinadas colectivamente. A diferencia de la opacidad propietaria de los LLM comerciales, el marco de *Wikimedia* garantiza historiales de versiones transparentes que permiten la trazabilidad clara de la procedencia de cualquier afirmación. Este compromiso con la agencia humana se plasma, además, en la estrategia de IA de la *Fundación* para 2025, que centra explícitamente los modelos de código abierto y la cobertura multilingüe equitativa con el fin de apoyar, y no de suplantar, el juicio editorial humano (Albon; Zia, 2025). Al resistir la sustitución del conocimiento curado por humanos con producciones automatizadas, esta postura de *prioridad humana* proporciona un modelo de gobernanza para un ecosistema de IA científica donde la comunidad conserva el control último sobre los instrumentos de su propia síntesis.

Quizá lo más significativo sea que *Wikimedia* está enfrentando actualmente la misma dinámica extractiva que amenaza al registro científico; mientras los sistemas comerciales de IA explotan de forma intensiva los datos de *Wikipedia* bajo licencia libre, simultáneamente desvían a los usuarios de la fuente original, contribuyendo a una disminución documentada del 8% en el número de visitantes en 2025 (*Wikimedia Foundation*, 2025). Este patrón, en el que un bien público generado por la comunidad es absorbido por productos privados que a su vez socavan la infraestructura originaria, es estructuralmente idéntico a la paradoja fiduciaria que enfrenta la investigación financiada con fondos públicos. Los experimentos en curso de la *Fundación* en materia de innovación institucional, incluyendo APIs empresariales y acuerdos de licencia sostenibles, representan así una base de evidencia esencial para la comunidad científica en su empeño por asegurar su propia soberanía computacional y epistémica. Mientras el *CERN* demuestra que la comunidad internacional puede construir y gobernar infraestructura física y digital compartida a gran escala, el modelo de *Wikimedia* demuestra que la gobernanza comunitaria del conocimiento es en sí misma factible, sostenible y capaz de producir resultados que las alternativas comerciales no han logrado replicar. Una IA científica de propiedad comunitaria se nutriría de ambos precedentes: el modelo del *CERN* para la infraestructura computacional compartida y la gobernanza entre naciones, y el modelo de *Wikimedia* para la curación del conocimiento gobernada por la comunidad, el versionado transparente y la integración de la IA fundamentada en principios.

8. Consideración de los contraargumentos

Si bien la fundamentación de la IA gobernada por la comunidad se apoya en líneas de evidencia convergentes, varias objeciones prominentes, relativas a la eficiencia de mercado, la velocidad institucional, la seguridad y la viabilidad fiscal, merecen una consideración detenida. Un contraargumento primario sugiere que la rápida mejora de la IA comercial, en particular la proliferación de modelos de «pesos abiertos» (*open-weight*) como *LLaMA* o *Mistral*, torna redundante la infraestructura pública. No obstante, esta perspectiva confunde acceso con gobernanza. Si bien los pesos abiertos permiten la ejecución local, no confieren a la comunidad científica agencia sobre el ciclo de vida del modelo. Las cuestiones fundamentales, la composición de los datos de entrenamiento, los objetivos de optimización, los calendarios de discontinuación, permanecen como prerrogativa exclusiva de los actores privados. Además, el incentivo estructural para que las corporaciones restrinjan la transparencia a medida que se intensifican las presiones competitivas, tendencia ya documentada en la decreciente apertura de los modelos de frontera (**Pei; Huang**, 2025), sugiere que confiar en la «generosidad» del código abierto comercial constituye un fundamento inestable para el registro científico.

Una segunda objeción se refiere a la supuesta inercia de la coordinación internacional, argumentando que los esfuerzos científicos multilaterales son demasiado lentos para seguir el ritmo de la frontera de la IA. Si bien las complejidades políticas de tales iniciativas son innegables, los precedentes históricos del *CERN*, el *Proyecto Genoma Humano* y la *Estación Espacial Internacional* demuestran que la coordinación científica de alta complejidad es factible cuando los riesgos epistémicos son suficientemente claros. Además, el panorama institucional contemporáneo está singularmente preparado para esta tarea. El surgimiento del *Panel Científico* de la *ONU* y los mecanismos

de implementación del *Pacto Digital Global* proporcionan un nivel de «andamiaje institucional» preexistente (*United Nations General Assembly, 2025; Independent International Scientific Panel on AI, s. f.*); que estuvo ausente en las generaciones anteriores de colaboraciones de la «Gran Ciencia» (*Big Science*).

En lo que respecta a las preocupaciones sobre seguridad y riesgos de doble uso, los críticos suelen argumentar que una infraestructura pública de IA podría facilitar el uso indebido de modelos potentes. Esta preocupación, aunque legítima, se aplica con igual o mayor fuerza a los sistemas propietarios que operan a puertas cerradas sin rendición de cuentas pública. Un marco gobernado por la comunidad permitiría la deliberación transparente y multilateral sobre las restricciones de uso y los protocolos de seguridad, en lugar de delegar decisiones éticas de tal envergadura a los equipos de gestión corporativa. Al alinearse con el mandato no militar del *Panel Científico* de la *ONU* (*United Nations General Assembly, 2025*), una IA científica de propiedad comunitaria podría establecer un precedente riguroso de delimitación de alcance.

Un marco gobernado por la comunidad permitiría la deliberación transparente y multilateral sobre las restricciones de uso y los protocolos de seguridad, en lugar de delegar decisiones éticas de tal envergadura a los equipos de gestión corporativa

El argumento de que los costes son prohibitivos queda cada vez más desmentido por el actual auge de las inversiones en «IA soberana» (*Innovation, Science and Economic Development Canada, 2025*). A medida que las naciones individuales comprometen miles de millones en capacidad computacional doméstica, resulta evidente que la comunidad investigadora global dispone del capital necesario. El desafío reside menos en la escasez de recursos que en la reorientación estratégica de la voluntad política, desplazándola desde la adquisición recurrente de licencias comerciales hacia la capitalización de infraestructura pública duradera. En última instancia, el coste de la inacción, medido en la pérdida de autonomía epistémica y la ampliación de la brecha global de computación, supera con creces los requisitos fiscales de un colectivo científico coordinado.

9. Más allá de la ciencia: la crisis de la información pública

El argumento a favor de una infraestructura de IA de propiedad comunitaria, si bien se fundamenta en las necesidades de la comunidad científica, se extiende con creces más allá del laboratorio. Los mismos LLM comerciales que median la indagación científica son ahora los instrumentos principales a través de los cuales cientos de millones de personas buscan respuestas sobre su salud, sus derechos y el mundo que les rodea. Los riesgos epistémicos descritos para la ciencia, alucinación, sesgo, opacidad, desajuste comercial, se amplifican cuando la población en general depende de estos sistemas para sus necesidades informativas cotidianas sin la formación metodológica ni las salvaguardas institucionales de que disponen en general la academia.

La evidencia proveniente de la salud pública es particularmente instructiva. Un estudio cualitativo que empleó protocolos de pensamiento en voz alta (*think-aloud*) encontró que los participantes integraban herramientas de IA como *ChatGPT* en sus rutinas de búsqueda de información sanitaria, valorándolas a menudo por su rapidez y claridad,

aun cuando albergaban dudas sobre su precisión, lo que revela una disociación entre confianza y uso (**Wardle; Urbani; Wang, 2025**). Los participantes utilizaron *ChatGPT* con frecuencia pese a preocupaciones explícitas sobre la procedencia y el sesgo, lo que sugiere que la conveniencia puede prevalecer sobre la cautela epistémica en la búsqueda de información cotidiana. Una evaluación translingüística de la información sobre vacunación proporcionada por *ChatGPT* encontró que, si bien las respuestas eran en buena medida precisas, los niveles de legibilidad superaban sistemáticamente los umbrales recomendados, en particular en inglés, y las respuestas en español parecían en ocasiones traducciones mecánicas que no contemplaban el contexto cultural (**Joshi et al., 2024**). Estos hallazgos son congruentes con la evidencia que muestra que los chatbots comerciales producen con frecuencia respuestas inexactas, generalizadas o sesgadas en la comunicación sanitaria, y que tales limitaciones afectan de manera desproporcionada a poblaciones vulnerables con alfabetización digital o sanitaria limitada (**Uddin, 2025**).

Las implicaciones se extienden a poblaciones sensibles. Un estudio que examinó las respuestas de chatbots de IA a consultas de adolescentes sobre alimentación, peso corporal y apariencia encontró que los consejos se enmarcaban con frecuencia en ideales sociales que podrían resultar perjudiciales para adolescentes con sintomatología de trastornos alimentarios, y que los chatbots no orientaban a los usuarios hacia recursos de apoyo regulados (**Sheen, 2025**). En farmacovigilancia, un dominio donde las imprecisiones acarrearán consecuencias directas para la seguridad del paciente, los investigadores han documentado la necesidad de salvaguardas diseñadas *ad hoc* para prevenir alucinaciones de los LLM que involucren nombres incorrectos de fármacos o términos de eventos adversos, precisamente porque el comportamiento por defecto de los modelos comerciales es inseguro para tales aplicaciones (**Hakim, 2025**). Una revisión rápida sobre el uso confiable y responsable de la IA generativa en contextos de salud pública identificó el sesgo algorítmico, el mal uso de datos y la ampliación de la desinformación sanitaria como riesgos que requieren mitigación a través de la supervisión humana, la transparencia y la comunicación culturalmente pertinente, ninguno de los cuales está garantizado estructuralmente por el modelo comercial (**Mackay; Kukan; McWhirter, 2025**). Lo que emerge de esta literatura es un patrón estructural: los sistemas de IA comerciales están optimizados para la interacción (*engagement*) y la utilidad general, no para el bienestar epistémico de las poblaciones que dependen cada vez más de ellos. El despliegue de modelos propietarios en contextos del Sur Global revela una profunda desconexión entre la optimización comercial y la necesidad local. Sin un control soberano sobre los datos de entrenamiento y las estructuras de rendición de cuentas, el uso de estas herramientas para la búsqueda de conocimiento esencial sigue siendo un acto de dependencia, donde el coste del error recae íntegramente sobre el usuario final.

Sin un control soberano sobre los datos de entrenamiento y las estructuras de rendición de cuentas, el uso de estas herramientas para la búsqueda de conocimiento esencial sigue siendo un acto de dependencia, donde el coste del error recae íntegramente sobre el usuario final

El modelo de *Wikipedia* descrito en la Sección 7 ofrece una poderosa contranarrativa. Durante más de dos décadas, *Wikipedia* ha proporcionado a la población mundial conocimiento enciclopédico de libre acceso, curado por la comunidad y con fuentes transparentes, regido no por el valor para los accionistas sino por estándares editoriales colectivos de verificabilidad, neutralidad y citación. Que las empresas de IA hayan consumido el contenido de *Wikipedia* como datos de entrenamiento, al tiempo que producen sistemas que desvían el tráfico de la propia enciclopedia, constituye una parábola de la dinámica extractiva más amplia que está en curso. La respuesta de la *Fundación Wikimedia*, redoblar la apuesta por una gobernanza centrada en el ser humano, herramientas de IA de código abierto y equidad multilingüe, en lugar de competir en términos comerciales, representa el tipo de postura institucional que el ecosistema informativo en su conjunto necesita.

Si bien resulta esencial para la integridad científica, la exigencia de una infraestructura de propiedad comunitaria se arraiga igualmente en la necesidad de alternativas sostenibles y de interés público frente a la extracción comercial de datos. Se trata de una cuestión de salud epistémica pública. Si los instrumentos a través de los cuales miles de millones de personas acceden y sintetizan el conocimiento son gobernados exclusivamente por entidades que maximizan el beneficio, el entorno informativo resultante reflejará incentivos comerciales en detrimento de las necesidades públicas. Una infraestructura de IA de propiedad comunitaria para la ciencia, si se diseña con la suficiente apertura, podría servir como núcleo de un ecosistema más amplio de herramientas de conocimiento gobernadas públicamente, uno que extienda el ethos de *Wikipedia*, conocimiento transparente y curado por la comunidad, a la era de la IA generativa.

10. Robustez de la confirmación: líneas de evidencia convergentes

La necesidad de una infraestructura de IA gobernada por la comunidad se sustenta en una convergencia robusta de evidencia que abarca los dominios empírico, estructural y normativo. Individualmente, estas líneas de indagación señalan vulnerabilidades específicas del *statu quo*; de manera colectiva, constituyen un argumento riguroso a favor de la transformación institucional. En el nivel fundacional, las evaluaciones empíricas de la IA comercial en contextos científicos han documentado de forma reiterada limitaciones significativas, incluyendo altas tasas de alucinación, la propagación de sesgos algorítmicos y fallos sistémicos en la reproducibilidad dentro de aplicaciones tanto clínicas como de investigación fundamental (**Mansoor et al., 2026; Ma et al., 2025; Limon et al., 2025; Kim; Choi, 2025**). Esta fragilidad empírica se ve agravada por una divergencia estructural cada vez más profunda entre la industria y la academia. Los análisis actuales indican que las entidades comerciales ejercen un dominio sin precedentes sobre el desarrollo de modelos, la captación de talento y los recursos computacionales necesarios, un desplazamiento que impide de facto a la comunidad académica contribuir a la producción de IA de frontera (**Pei; Huang, 2025; Ahmed; Wahed; Thompson, 2023**).

Las consecuencias de esta asimetría son particularmente pronunciadas en lo que concierne a la equidad en la investigación global. La evidencia cuantitativa de los estudios sobre datos abiertos demuestra que la infraestructura accesible incrementa de forma mensurable la participación de investigadores de regiones subrepresentadas

(**Charpignon et al.**, 2024), lo que sugiere que una infraestructura de IA abierta y soberana ejercería un efecto democratizador análogo. A la inversa, los riesgos de mantener el modelo actual, mediado comercialmente, quedan subrayados por la literatura de salud pública que documenta daños epistémicos a nivel poblacional. Estos estudios revelan que los sistemas de IA propietarios producen con frecuencia información sanitaria sesgada, inconsistente o inaccesible para las poblaciones vulnerables, lo que indica que los riesgos de la gobernanza privada se extienden más allá del laboratorio hacia el ecosistema global de información (**Wardle; Urbani; Wang**, 2025; **Joshi et al.**, 2024; **Uddin**, 2025; **Sheen et al.**, 2025; **Hakim et al.**, 2025; **MacKay; Kukan; McWhirter**, 2025).

Pese a estos desafíos, se perfila un camino claro gracias a los recientes avances normativos e institucionales. Un creciente consenso multilateral se evidencia en el *Pacto Digital Global*, que enmarca explícitamente los modelos de IA abierta como bienes públicos digitales (*United Nations General Assembly*, 2025). Este compromiso político se está materializando a través de nuevos mecanismos institucionales, como el *Panel Científico Internacional Independiente sobre IA* y el *Diálogo Global sobre Gobernanza de la IA* (*Independent International Scientific Panel on AI*, s. f.; **Guterres**, 2026). Además, la viabilidad de esta transición se ancla en precedentes institucionales consolidados. El éxito operativo sostenido del *CERN* y otras infraestructuras de investigación gobernadas por la comunidad, junto con el modelo de la *Fundación Wikimedia* para la curación del conocimiento basada en principios, proporciona un modelo replicable para la gobernanza a gran escala en beneficio del interés público. Estos precedentes, combinados con el auge de las iniciativas nacionales de IA soberana, validan la viabilidad técnica y organizativa de la infraestructura pública a escala. La defensa de la soberanía computacional trasciende los ajustes técnicos puntuales para representar una evolución institucional fundamental, necesaria para salvaguardar la integridad y la equidad del registro científico global.

11. Un camino a seguir: del diagnóstico a la acción

La comunidad científica global se encuentra en una encrucijada crítica en la evolución de la producción de conocimiento. No estamos confinados a una elección binaria entre una «brecha digital» caracterizada por la exclusión total y un «clientelismo digital» en el que el Sur Global permanece perpetuamente dependiente de infraestructuras propietarias. Existe un tercer camino viable: el establecimiento de una Infraestructura Pública Digital para la Ciencia, de propiedad comunitaria, gobernada por pares y fundamentada en el principio de que la síntesis del conocimiento humano es un derecho universal y no un producto comercial. La arquitectura institucional para respaldar esta vía está ahora al alcance, pero requiere una acción coordinada en varios frentes.

El *ISC* es el organismo mejor posicionado para sentar las bases técnicas. Hacemos un llamamiento al *ISC* para que convoque un grupo de trabajo global especializado encargado de elaborar las especificaciones técnicas y los marcos de gobernanza para una «Red Global de

No estamos confinados a una elección binaria entre una «brecha digital» caracterizada por la exclusión total y un «clientelismo digital» en el que el Sur Global permanece perpetuamente dependiente de infraestructuras propietarias

Computación Científica para AI». Este organismo debe superar el modo consultivo que ha caracterizado los esfuerzos precedentes: ha de integrar la experiencia de investigadores en IA, científicos de dominio, profesionales de la información y especialistas en integridad de la investigación, con una representación robusta de los PLEMEDS. Iniciativas como *Latam-GPT* demuestran que el capital humano y la voluntad institucional para dicha colaboración ya existen en el Sur Global; lo que falta es la infraestructura de coordinación.

Los fundamentos normativos de dicha infraestructura deben consolidarse simultáneamente en el seno de los estándares internacionales. La *Unesco* debería ampliar su *Recomendación sobre Ciencia Abierta* para reconocer explícitamente la infraestructura soberana de IA como un componente necesario de la capacidad científica. El *Pacto Digital Global* proporciona el mandato político de alto nivel; la *Unesco* se encuentra en una posición privilegiada para aportar la especificidad operativa, las directrices de implementación, los mecanismos de seguimiento y los criterios de evaluación comparativa, que los Estados miembros necesitan para traducir los principios en política nacional.

La sostenibilidad financiera exige una estrategia de financiación coordinada que rompa con el modelo actual de gasto comercial disperso. El *Consejo Global de Investigación (GRC)*, por sus siglas en inglés) debería establecer una dotación computacional compartida (*shared computational endowment*), inspirada en los acuerdos de financiación que han sostenido al *CERN* durante décadas y que se están adaptando ahora para los programas nacionales de computación soberana de IA en Canadá, la Unión Europea y otros lugares. Al aunar las contribuciones nacionales y contrastarlas con los miles de millones que los Estados miembros ya invierten en licencias de software comercial, suscripciones a bases de datos propietarias y cargos por procesamiento de artículos, el *GRC* puede movilizar el capital necesario para construir un activo público duradero en lugar de perpetuar un ciclo de extracción de rentas. El recién creado *Panel Científico Internacional Independiente sobre IA* debe emplear su mandato para abordar la dimensión estructural del problema. Un examen dedicado de los riesgos que la concentración comercial de la IA plantea a la integridad científica y a la equidad epistémica, así como del potencial de las alternativas de propiedad comunitaria para mitigar dichos riesgos, debe figurar en la evaluación inaugural del *Panel*. El mandato del *Panel* para evaluar «oportunidades, riesgos e impactos» de la IA comprende esta cuestión; lo que se necesita es que la comunidad científica articule la argumentación con el rigor suficiente para inscribirla en la agenda.

Estos desplazamientos institucionales no prosperarán, sin embargo, sin la defensa sostenida por parte de los investigadores individuales. Este es también un llamamiento a los científicos para que documenten de manera sistemática los casos en que las limitaciones de la IA comercial, opacidad propietaria, sesgo algorítmico, producciones alucinadas, exclusión lingüística, hayan comprometido la calidad de su investigación o exacerbado la inequidad. La experiencia de la comunidad de *Wikimedia* al crear un *WikiProyecto* dedicado a identificar y eliminar la desinformación generada por IA en *Wikipedia* ofrece un modelo para este tipo de aseguramiento distribuido de la calidad, impulsado por la comunidad. Dicha documentación proporciona la base probatoria sin la cual la acción institucional permanece en el terreno de la aspiración.

La ventana para establecer la gobernanza comunitaria sobre los instrumentos de síntesis del conocimiento se está estrechando. A medida que los sistemas propietarios se arraigan más profundamente en los flujos de trabajo de investigación, las dependencias de trayectoria (*path dependencies*) harán que las alternativas sean cada vez más difíciles de implementar. La arquitectura multilateral que se está construyendo, el *Pacto Digital Global*, el *Panel Científico*, el *Diálogo Global*, representan una oportunidad generacional para integrar el principio de propiedad comunitaria en los fundamentos digitales de la ciencia. Ha llegado el momento de actuar con decisión.

12. Referencias

Ahmed, N.; Wahed, M.; Thompson, N. C. (2023). The growing influence of industry in AI research. *Science*, 379(6635):884–886.

<https://doi.org/10.1126/science.ade2420>

Albon, C.; Zia, L. (2025). Artificial intelligence for editors: strategy brief. *Wikimedia Foundation*, April

https://meta.wikimedia.org/wiki/Strategy/Multigenerational/Artificial_intelligence_for_editors

Asgari, E.; Montaña-Brown, N.; Dubois, M.; Khalil, S.; Balloch, J.; Yeung, J. A.; Pimenta, D. (2025). A framework to assess clinical safety and hallucination rates of LLMs for medical text summarisation. *NPJ digital medicine*, 8(1), 274.

<https://doi.org/10.1038/s41746-025-01670-7>

Bommasani, R.; Hudson, D. A.; Adeli, E.; Altman, R.; Arora, S.; Von Arx, S.; ...Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

<https://doi.org/10.48550/arXiv.2108.07258>

Bona, M.; Schopper, H.; Spiro, M. (2016). Science: a model for collaboration? *CERN Courier*.

<https://CERNcourier.com/a/viewpoint-science-a-model-for-collaboration>

CERN Open Science Office. Open infrastructure.

<https://openscience.CERN/infrastructure>

Charpignon M. L.; Celi, L. A.; Cobanaj, M.; Eber, R.; Fiske, A.; Gallifant, J.; Li, C.; Lingamallu, G.; Petushkov, A.; Pierce, R. (2024). Diversity and inclusion: a hidden additional benefit of Open Data. *PLOS Digit Health*, 3(7):e0000486.

<https://doi.org/10.1371/journal.pdig.0000486>

Chelli, M.; Descamps, J.; Lavoué, V.; Trojani, C.; Azar, M.; Deckert, M.; Raynier, J. L.; Clowez, G.; Boileau, P.; Ruetsch-Chelli, C. (2024). Hallucination rates and reference accuracy of ChatGPT and bard for systematic reviews: comparative analysis. *Journal of Medical Internet Research*, 26(1), e53164.

<https://doi.org/10.2196/53164>

Etchemendy, J.; Landay, J.; Li, F. F.; Manning, C. (October 30, 2025). *Universities must reclaim AI research for the public good*. Stanford Institute for Human-Centered Artificial Intelligence.

<https://hai.stanford.edu/news/universities-must-reclaim-ai-research-for-the-public-good>

Guterres, A. (2026). Remarks to the first meeting of the Independent International Scientific Panel on Artificial Intelligence. United Nations Secretary-General.

<https://www.un.org/sg/en/content/sg/statements/2026-03-03/un-secretary-generals-remarks-the-first-meeting-of-the-independent-international-scientific-panel-artificial-intelligence-delivered>

Hakim, J. B.; Painter, J. L.; Ramcharran, D.; Kara, V.; Powell, G.; Sobczak, P.; Sato, C.; Bate, A.; Beam, A. (2025). The need for guardrails with large language models in pharmacovigilance and other medical safety critical settings. *Sci Rep.*, 15(1):27886.

<https://doi.org/10.1038/s41598-025-09138-0>

Hedderich, M. A.; Lange, L.; Adel, H.; Strötgen, J.; Klakow, D. (2021). A survey on recent approaches for natural language processing in low-resource scenarios. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2021:2545–2568.

<https://doi.org/10.18653/v1/2021.naacl-main.201>

Heracleous, L.; Gößwein, J.; Beaudette, P. (2018). Open strategy-making at the Wikimedia foundation: a dialogic perspective. *The Journal of Applied Behavioral Science*, 54(1), 5-35.

<https://doi.org/10.1177/0021886317712665>

Hosseini, M.; Horbach, S. P. J. M.; Holmes, K.; Ross-Hellauer, T. (2025). Open science at the generative AI turn: an exploratory analysis of challenges and opportunities. *Quantitative Science Studies*, 6: 22–45.

https://doi.org/10.1162/qss_a_00337

Innovation, Science and Economic Development Canada (2025). *Canadian Sovereign AI Compute Strategy*. Government of Canada.

<https://ised-isde.canada.ca/site/ised/en/canadian-sovereign-ai-compute-strategy>

Joshi, S.; Ha, E.; Amaya, A.; Mendoza, M.; Rivera, Y.; Singh, V. K. (2024). Ensuring accuracy and equity in vaccination information from ChatGPT and CDC: mixed-methods cross-language evaluation. *JMIR Formative Research*, 8:e60939.

<https://doi.org/10.2196/60939>

Kaushik, A.; Barcellona, C.; Mandyam, N. K.; Tan, S. Y.; Tromp, J. (2025). Challenges and opportunities for data sharing related to artificial intelligence tools in health care in low- and middle-income countries: systematic review and case study from Thailand. *Journal of Medical Internet Research*, 27:e58338.

<https://doi.org/10.2196/58338>

Kim, D.; Choi, J. (2025). Big data and AI: potential and challenges for digital transformation in toxicology. *Environ Anal Health Toxicology*, 40:e2025s07.
<https://doi.org/10.5620/eaht.2025s07>

Latam-GPT. Frequently asked questions.
<https://www.latamgpt.org/en/faq>

Levy-Yeyati, E. (2025). *Latam-GPT and the search for AI sovereignty*. Brookings Institution, November 25.
<https://www.brookings.edu/articles/latam-gpt-and-the-search-for-ai-sovereignty>

Limon, D.; Satish, V.; Raghavan, N.; Morris, M. X.; Muir, M. T.; Rajesh, A. (2025). Artificial intelligence in surgery revisited: a 2025 update on machine learning for predicting complications and outcomes. *The American Surgeon*, 92(3):658-674.
<https://doi.org/10.1177/00031348251393934>

Ma, Z.; Liu, Y.; Zhang, Z.; Chen, R.; Fan, H.; Cao, X.; Ni, L. (2025). Clinical applications of large language models in knee osteoarthritis: a systematic review. *Front Med*, 12:1670824.
<https://doi.org/10.3389/fmed.2025.1670824>

Mackay, M.; Kukan, A.; McWhirter, J. E. (2025). The double-edged algorithm: a rapid review exploring the trustworthy and responsible use of generative AI in public health. *Digital Health*, 11:20552076251393302.
<https://doi.org/10.1177/20552076251393302>

Mann, T. (2025). Sorry, but DeepSeek didn't really train its flagship model for \$294,000. *The Register*. September 19.
https://www.theregister.com/2025/09/19/deepseek_cost_train

Mansoor, I.; Abdullah, M.; Rizwan, M. D.; Fraz, M. M. (2026). Reasoning with large language models in medicine: a systematic review of techniques, challenges and clinical integration. *Health Information Science and Systems*, 14(1), 6.
<https://doi.org/10.1007/s13755-025-00403-0>

Mboa-Nkoudou, T. H. (2023). We need a decolonized appropriation of AI in Africa. *Nature Human Behaviour*, 7:1810–1811.
<https://doi.org/10.1038/s41562-023-01741-3>

Moodley, K.; Malpani, R.; Reis, A. A. (2025). How will 'Chat-IRB' impact research ethics review in LMICs? *Journal of Medical Ethics*.
<https://doi.org/10.1136/jme-2025-111452>

Omar, M.; Vera, S.; Collins, J. D.; Reich, D.; Freeman, R.; Gavin, N.; Charney, A.; Stump, L.; Bragazzi, N. L.; Nadkarni, G. N.; Klang, E. (2025). Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support. *Communications Medicine*, 5(1), 330.
<https://doi.org/10.1038/s43856-025-01021-3>

Pei, G.; Huang, H. (2025). Open science falling behind in the era of artificial intelligence. *Front Res Metr Anal*, 10:1595824.
<https://doi.org/10.3389/frma.2025.1595824>

Sheen, F.; Mullarkey, B.; Witcomb, G. L.; Opitz, M. C.; Maloney, E.; Baldoza, S. M.; White, H. J. (2025). How do artificial intelligence chatbots respond to questions from adolescent personas about their eating, body weight or appearance? *Child and Adolescent Mental Health*.
<https://doi.org/10.1111/camh.70047>

Shool, S.; Adimi, S.; Saboori-Amleshi, R.; Bitaraf, E.; Golpira, R.; Tara, M. (2025). A systematic review of large language model (LLM) evaluations in clinical medicine. *BMC Medical Informatics and Decision Making*, 25(1), 117.
<https://doi.org/10.1186/s12911-025-02954-4>

Shyr, C.; Ren, B.; Hsu, C. Y.; Tinker, R. J.; Cassini, T. A.; Hamid, R.; Wright, A.; Bastarache, L.; Peterson, J. F.; Malin, B. A.; Xu, H. (2025). A statistical framework for evaluating the repeatability and reproducibility of large language models. *medRxiv*, 2025-08.
<https://doi.org/10.1101/2025.08.06.25333170>

Siefert-McCanse, D. (2025). *NAIRR Pilot inaugural annual meeting final report*.
<https://par.nsf.gov/biblio/10621635>

Sun, Z.; Yim, W. W.; Uzuner, Ö.; Xia, F.; Yetisgen, M. (2025). A scoping review of natural language processing in addressing medically inaccurate information: Errors, misinformation, and hallucination. *Journal of Biomedical Informatics*, 104866.
<https://doi.org/10.1016/j.jbi.2025.104866>

Tran, V.; Lopez, A.; Nguyen, G.; Sáinz-Pardo, J.; Moltó, G. (2025). AI4EOSC: Artificial Intelligence for the European Open Science Cloud. In: *Academia Sinica Grid Computing Centre (ASGC). International Symposium on Grids and Clouds (ISGC2025)* (Vol. 16, p. 21), March.
<https://pos.sissa.it/488/003/pdf>

Uddin, J. (2025). Conversational AI in healthcare communication: opportunities, risks, and implications for health equity. *Journal of Communication in Healthcare*, 1–4.
<https://doi.org/10.1080/17538068.2025.2594768>

Unesco (2021). Recommendation on the Ethics of Artificial Intelligence. Adopted by the General Conference at its 41st session.
<https://unesdoc.unesco.org/ark:/48223/pf0000380455>

Unesco (2024). Exploring opportunities and challenges at the intersection of open science and artificial intelligence.
<https://www.unesco.org/en/articles/exploring-opportunities-and-challenges-intersection-open-science-and-artificial-intelligence>

United Nations (2025). Independent International Scientific Panel on AI. Frequently asked questions.

<https://www.un.org/independent-international-scientific-panel-ai/en/faq>

United Nations General Assembly (2024). The pact for the future. Resolution A/RES/79/1. 22, September.

<https://docs.un.org/en/a/res/79/1>

United Nations General Assembly (2025). Terms of reference and modalities for the establishment and functioning of the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on Artificial Intelligence Governance. Resolution A/RES/79/325. August, 26th.

<https://docs.un.org/en/A/RES/79/325>

Vedder, A.; Nisevic, M.; Fleerackers, S. (2025). Ethics first? On the EU approach to AI governance. In: Raposo, Vera Lúcia (ed.). *The European Artificial Intelligence Act: Promises and Perils?* (pp. 3-21). Cham: Springer Nature Switzerland.

https://doi.org/10.1007/978-3-031-98406-8_1

Voss, J. (2005). Measuring *Wikipedia*. In: Ingwersen, P.; Larsen, B. (eds.). *Proceedings of ISSI*, vol. 1, pp. 221-231, July.

https://www.issi-society.org/proceedings/issi_2005/Voss_ISSI2005.pdf

Wang, G.; Zhang, K.; Jiang, J.; Wang, C.; Bi, H.; Liang, H.; Qi, Z.; Huang, Y.; Li, Y.; Yang, X. (2026). Human–large language model collaboration in clinical medicine: a systematic review and meta-analysis. *NPJ Digital Medicine*, 9, 195.

<https://doi.org/10.1038/s41746-026-02382-2>

Wardle, C.; Urbani, S.; Wang, E. (2025). Evolving health information-seeking behavior in the context of Google AI Overviews, ChatGPT, and Alexa: interview study using the think-aloud protocol. *Journal of Medical Internet Research*, 27:e79961.

<https://doi.org/10.2196/79961>

Wikimedia Foundation (2025). Our new AI strategy puts *Wikipedia's* humans first, April 30.

<https://Wikimediafoundation.org/news/2025/04/30/our-new-ai-strategy-puts-Wikipedias-humans-first>