

Cuando la respuesta ya habla sola

Fabrizio Gaviña

Licenciado en Bibliotecología y Documentación · Argentina

<https://orcid.org/0009-0007-1776-791X>

gavinafabrizio@gmail.com

Resumen

El acceso abierto latinoamericano fue pensado como política de reciprocidad. Que el conocimiento financiado colectivamente circule dentro de la comunidad que lo produce. Esa premisa sigue intacta en el texto de las normas. Lo que cambió es el tipo de sistema con el que esas normas entran en contacto. Cuando los repositorios institucionales se convierten en corpus de entrenamiento para modelos de lenguaje privados, el mismo gesto que democratizaba la lectura habilita la extracción algorítmica. Articulando debates sobre colonialidad del saber, economía política de la información y bibliotecología crítica latinoamericana, el artículo sostiene que abrir no garantiza soberanía sobre lo que después se hace con lo abierto. El problema no es la apertura. El problema es quién transforma, a qué escala, bajo qué condiciones y a quién no le vuelve nada.

Palabras clave: acceso abierto · extracción algorítmica · soberanía epistémica · colonialidad del saber · infraestructura digital · inteligencia artificial generativa

Abstract

Latin American open access was conceived as a policy of reciprocity. Knowledge produced with collective resources was expected to circulate within the community that produced it. That premise survives intact in normative texts. What changed is the type of system those norms now come into contact with. When institutional repositories become training corpora for private language models, the same gesture that democratized reading also enables algorithmic extraction. Bringing together debates on the coloniality of knowledge, the political economy of information, and Latin American critical

library science, this article argues that openness does not guarantee sovereignty over what is later done with openly available knowledge. The problem is not openness itself. The problem is who transforms knowledge, at what scale, under what conditions, and who receives nothing in return.

Keywords: open access · algorithmic extraction · epistemic sovereignty · coloniality of knowledge · digital infrastructure · generative artificial intelligence

Apertura sin equivalencia

Los repositorios institucionales latinoamericanos funcionan. Reciben producción científica, la organizan, la preservan, la hacen accesible. Nadie puede señalarles una falla técnica. La Ley 26.899 institucionalizó en 2013, después de años de discusión, la obligación de depositar en repositorios públicos la producción financiada con fondos del Estado. SciELO y CLACSO venían construyendo esa lógica desde los noventa, con la premisa de que el conocimiento producido en la región debía circular dentro de la región, en condiciones de reciprocidad académica y no de dependencia editorial respecto de los centros hegemónicos. Era una posición, no una política neutral. Implicaba que el circuito del saber todavía conservaba alguna relación con la comunidad que lo producía.

Eso sigue funcionando. El ecosistema para el que fue pensado, no.

Lo que ocurre ahora es esto. Los textos depositados son recolectados por rastreadores que construyen corpus masivos de entrenamiento para sistemas de lenguaje a gran escala. El protocolo OAI-PMH, que la comunidad bibliotecaria diseñó para garantizar interoperabilidad entre repositorios, expone los mismos registros de metadatos que ahora sirven para identificar y capturar esos contenidos a escala. Lo que esa ley organizó, además de acceso, fue disponibilidad técnica. La interoperabilidad académica mutó en interoperabilidad extractiva sin que ningún reglamento lo dispusiera ni ninguna institución lo discutiera. Los textos completos disponibles bajo licencias abiertas se descargan sin negociación institucional ni reconocimiento explícito. El artículo sigue siendo un artículo. La política de

acceso sigue siendo una política de acceso. Lo que cambió es lo que le ocurre al texto cuando empieza a circular.

En la lectura académica convencional, el texto tiene autor, contexto de producción, aparato crítico, posición dentro de un debate. En los sistemas de IA generativa, el texto deja de operar de esa manera. Se descompone en tokens, se integra en regularidades estadísticas, pierde singularidad frente al resto del corpus procesado. El sistema no preserva documentos. Reconstruye continuidad probabilística sobre lo que estuvo disponible con frecuencia suficiente. Lo que conserva singularidad documental tiende a tratarse como ruido. Bender et al. (2021) señalaron que esa operación produce consecuencias epistémicas invisibles desde la interfaz de uso. El sistema genera texto coherente sin que esa coherencia garantice una relación verificable con los materiales de los que partió.

Las políticas de acceso abierto en América Latina fueron construidas para otro problema. Paywalls, licencias exclusivas, concentración editorial en el norte global. Los regímenes de circulación del conocimiento no son técnicamente neutrales, señaló Vercelli (2009). Habilitan ciertos usos y restringen otros según las condiciones bajo las cuales fueron diseñados. Esas condiciones cambiaron. Las políticas, no. La apertura sigue ahí. La reciprocidad que la justificaba, no.

Cuando las condiciones de reutilización no están definidas con precisión, la ambigüedad favorece a quienes tienen mayor capacidad técnica y jurídica para interpretarla. Busaniche (2010) ya lo advertía. Los proyectos normativos en discusión en Argentina no contemplan mecanismos de protección frente a la reutilización del conocimiento depositado como corpus de entrenamiento (Vercelli, 2024). La inserción de los repositorios abiertos en el circuito de la IA generativa no rompe las reglas existentes. Funciona exactamente dentro de ellas.

La infraestructura que el discurso no nombra

El debate sobre el acceso abierto se formula casi siempre en términos de política de información, normativa institucional y práctica editorial. Ese

lenguaje nombra los problemas dentro del campo de lo administrable. Deja fuera la dimensión material. El acceso no existe sin infraestructura, y esa infraestructura no está distribuida simétricamente entre quienes producen el conocimiento y quienes lo procesan a escala global.

La pregunta por la circulación de la información es siempre también una pregunta por la propiedad y el control de los sistemas a través de los cuales esa circulación ocurre. Becerra y Mastrini (2009) lo mostraron al analizar la concentración mediática en América Latina. La concentración no necesita prohibir el acceso para estructurar la desigualdad; le basta con controlar la infraestructura de transformación. El argumento funciona igual en el ecosistema de la IA.

Los sistemas de inteligencia artificial generativa requieren una capacidad de cómputo que no está distribuida equitativamente. Los centros de datos donde se entrenan los modelos de lenguaje concentran recursos energéticos, chips especializados y arquitecturas técnicas cuya producción y mantenimiento dependen de cadenas de suministro controladas por un número reducido de actores. América Latina no participa en esa capacidad como productora sino como usuaria. La concentración del desarrollo de IA en pocas plataformas transnacionales reproduce las desigualdades del mercado cultural global. La capacidad de transformar el conocimiento no coincide con la capacidad de producirlo (Kulesz, 2018). El flujo de datos va de la periferia al centro. El flujo de retorno no opera bajo las mismas condiciones de reciprocidad.

América Latina abre su producción científica y la expone mediante protocolos de interoperabilidad diseñados para facilitar el acceso. Esa apertura alimenta sistemas que procesan el conocimiento regional e integran sus resultados en modelos globales que no reconocen su origen ni redistribuyen valor equivalente. En ese circuito, la apertura ya no democratiza; expone.

La noción de extractivismo tiene en América Latina un largo recorrido analítico aplicado a los recursos naturales y a las industrias culturales. En este contexto se vuelve pertinente con las modulaciones propias de la economía del conocimiento. No se trata de expropiación en sentido clásico ni de apropiación ilegal, sino de un uso intensivo de producción abierta dentro

de sistemas donde el retorno no está garantizado. El conocimiento producido localmente, con recursos públicos, bajo políticas construidas como alternativa a la dependencia editorial, termina funcionando como insumo de sistemas globales de generación de texto que no operan bajo las reglas de reciprocidad que lo hicieron posible.

Colonialidad del corpus como saber sin densidad estadística

La colonialidad del saber opera clasificando los conocimientos según su posición relativa en una geopolítica donde el norte global define los criterios de lo que cuenta como ciencia, como teoría, como referencia válida. Quijano (2000), Mignolo (2010) y Lander (2000) lo formularon de maneras distintas, pero el argumento converge. Las ciencias sociales latinoamericanas trabajan sobre categorías construidas en condiciones históricas ajenas a la región, que se presentan como universales.

Esa geopolítica del conocimiento encuentra en los sistemas de IA generativa una nueva formalización técnica. Los modelos aprenden de lo que está disponible con densidad estadística suficiente en los corpus de entrenamiento. La composición de esos corpus no es neutral. Más del noventa por ciento del corpus C4, uno de los conjuntos de entrenamiento más utilizados, estaba en inglés; el español apareció de manera marginal pese a ser la segunda lengua más hablada del planeta (Dodge et al., 2021). Los repositorios latinoamericanos tuvieron una representación menor, no porque la producción regional sea de menor calidad académica, sino porque esa producción no tuvo la densidad estadística necesaria en el rastreo web masivo sobre el que se construyeron los corpus.

Los sistemas de evaluación científica transnacional ya operaban como filtros, determinando qué producción periférica adquiere visibilidad internacional y en qué condiciones de subordinación conceptual (Beigel, 2017). Los investigadores de la región publican en inglés, adaptan sus marcos teóricos a los estándares de las revistas indexadas en el norte global, producen conocimiento que circula bajo condiciones de reconocimiento asimétrico. Los sistemas de IA generativa heredan esa asimetría, la amplían y la normalizan.

Lo que Ricaurte (2022) denominó violencia epistémica a escala opera aquí de manera específica. Los modelos producen activamente borraduras al generar texto que presenta como universal lo que es regional, como neutro lo que es situado, como conocimiento lo que es residuo estadístico de la concentración hegemónica.

En los sistemas de generación automática, el criterio de qué cuenta como conocimiento no depende de la validación disciplinar ni de la trayectoria institucional de quien lo produjo. Depende de la frecuencia de aparición en el corpus disponible. La exclusión no requiere acto deliberado; requiere ausencia de densidad estadística. Y esa ausencia no es técnica. Es el resultado de décadas de visibilidad desigual, de sistemas de indexación que favorecieron ciertos circuitos y descartaron otros. La colonialidad del conocimiento, en ese sentido, no quedó en los libros. Está en los pesos del modelo. Y lo que no alcanzó ese umbral de frecuencia, lo que el rastreo pasó por alto, lo que circuló de manera irregular o escasamente indexada, sigue existiendo fuera de esos pesos. Ahí sobrevive parte de la memoria bibliográfica regional.

La producción científica periférica funciona como participación asimétrica en el sistema global de ciencia. Los países del sur producen datos e insumos empíricos que se incorporan en marcos teóricos elaborados en el centro (Kreimer, 2010). Los sistemas de IA no necesitan voluntad extractiva para reproducir esa estructura. Les basta con aprender de lo disponible.

La transformación silenciosa del estatuto del conocimiento abierto

El acceso abierto fue diseñado para quien lee, interpreta, cita, contesta, acumula. La apertura produce circulación académica porque conecta textos con lectores que operan dentro de marcos de validación compartidos, donde el origen del texto importa, donde la autoría es auditable, donde la referencia garantiza un contrato de trazabilidad entre ideas y sus condiciones de producción. La bibliotecología latinoamericana construyó su práctica sobre ese contrato. La mediación, la catalogación, la referencia y la preservación tienen sentido como funciones dentro de ese circuito.

Los sistemas de inteligencia artificial generativa no son usuarios de ese circuito. Procesan el texto sin habitarlo. Donde el lector académico activa un aparato crítico para evaluar lo que lee, el sistema de IA lo descompone en unidades que alimentan un modelo probabilístico. La trazabilidad que la referencia garantizaba no tiene traducción en esa arquitectura. La bibliografía pierde espesor operativo cuando la procedencia deja de importar para el sistema que la procesa. El contexto de producción no se preserva. La autoría es irrelevante.

Esa transformación cambia el estatuto del acceso abierto como política. Si lo que se llama acceso implica también la posibilidad de que el texto sea procesado por sistemas que lo reconfiguren sin preservar su procedencia, entonces el acceso abierto ya no es únicamente una política de democratización de la lectura. El mismo gesto que habilita circulación académica habilita extracción algorítmica. Y esa doble condición no está inscrita en el lenguaje normativo que regula los repositorios. Nadie lo inscribió porque nadie lo previó. Y ahora el lenguaje institucional del acceso abierto sigue operando como si la extracción no estuviera ocurriendo.

El resultado no es que el conocimiento desaparezca. Cambia de lugar en el circuito. Esa disolución no es visible como pérdida porque el conocimiento sigue apareciendo en forma de respuestas, síntesis y enunciados plausibles. Lo que desaparece es la relación verificable entre ese resultado y su origen; y con ella, la posibilidad de debatir, atribuir, acumular que la referencia académica garantizaba como práctica colectiva.

Hay un punto en que el acceso deja de ser solo una política de circulación. No en el sentido jurídico ni institucional, sino en uno más silencioso. Lo que no está en circuitos accesibles no entra en lo que puede ser procesado, reorganizado o devuelto como conocimiento legible. Y lo que entra lo hace bajo reglas que no dependen de su contexto original de producción. El criterio de qué cuenta como conocimiento se desplaza, sin anuncio y sin decisión explícita, desde la validación disciplinar hacia la frecuencia de aparición en el corpus disponible. La apertura deja de ser democratización y empieza a operar como filtro de lo que podrá, después, decirse.

Soberanía del conocimiento en la era de la extracción distribuida

La soberanía informacional ha tenido en América Latina un recorrido específico, vinculado a debates sobre dependencia tecnológica, concentración mediática y control sobre los flujos de información en relaciones internacionales profundamente desiguales. Ese recorrido vuelve, pero desplazado. La soberanía del conocimiento ya no se juega en el control del acceso sino en las condiciones bajo las que ese conocimiento puede ser transformado a escala.

Tener el conocimiento disponible no garantiza control sobre lo que ocurre con él una vez que entra en sistemas de procesamiento global. Los repositorios están abiertos y la producción es accesible. Todo funciona. Y sin embargo el conocimiento producido localmente alimenta modelos globales cuya gobernanza no incluye representación latinoamericana, cuyos resultados no reconocen su origen regional, y cuyos beneficios no se redistribuyen en términos equivalentes a los productores que los alimentaron.

La participación de la bibliotecología latinoamericana en los espacios donde se define la gobernanza de la IA sigue siendo marginal (D'Amico y Vercelli, 2024). Esa marginalidad responde a una posición estructural. Las decisiones sobre plataformas, estándares y condiciones de uso del conocimiento abierto se toman en espacios donde la región tiene escasa incidencia. La UNESCO aprobó en 2021 una Recomendación sobre la ética de la inteligencia artificial que enfatiza la necesidad de preservar la diversidad cultural y lingüística en el desarrollo de estos sistemas. Que ese marco normativo exista no equivale a que opere. La brecha entre enunciado normativo e incidencia real es donde la soberanía deja de ser principio.

Resistir la violencia epistémica a escala requiere construir alternativas reales, no solo formular críticas al estado actual (Ricaurte, 2022). Cuál es el margen de acción disponible cuando la capacidad de transformar el conocimiento está concentrada fuera de la región; cuáles son las condiciones bajo las cuales los repositorios podrían definir términos de reuso que trasciendan el acceso; qué implica para la mediación bibliotecaria el hecho de que su objeto ya no sea únicamente el encuentro entre un lector y un texto sino también

quién procesa ese texto y bajo qué reglas. Esas preguntas no tienen respuesta fácil.

Los imaginarios sociotécnicos que orientan la adopción de sistemas de IA en la región no se construyen localmente sino que se importan junto con los sistemas mismos, reproduciendo formas de dependencia que la región conoce bien bajo otras denominaciones (Ricaurte, Gómez-Cruz y Siles, 2024). La apertura, en ese marco, garantiza disponibilidad. Lo que no garantiza es el camino de vuelta.

Una disputa que el lenguaje del acceso abierto no puede nombrar

El acceso abierto no está en crisis porque haya fallado como política. Está en tensión porque el circuito en que opera ya no es el que le dio sentido. Lo que fue concebido como democratización de la circulación opera también, en las condiciones actuales, como ampliación de la superficie disponible para la extracción algorítmica. Esas dos funciones no se cancelan mutuamente, pero tampoco son equivalentes. Y la diferencia entre ellas define el problema político que la bibliotecología latinoamericana tiene frente a sí.

Sostener que el acceso abierto sigue siendo una política válida es correcto. Sostener que esa validez es suficiente sin revisar las condiciones en que opera actualmente es una forma de neutralidad que el problema no admite. La apertura fue pensada para otro tipo de problema. El terreno donde aterrizó la usa para otra cosa. El conocimiento producido en América Latina, abierto para circulación académica, disponible para reciprocidad disciplinar, entra también en sistemas que lo procesan sin preservar su trazabilidad, sin reconocer su origen, sin devolver valor en términos equivalentes a quienes lo produjeron.

La respuesta a esa situación no está disponible dentro de las categorías con las que la bibliotecología latinoamericana pensó históricamente su práctica. La pregunta ya no es si abrir o no abrir. La pregunta es qué tiene que cambiar para que la apertura produzca circulación y no solo disponibilidad; para que la soberanía del conocimiento no quede reducida al momento de la publicación

mientras se pierde en todos los procesos que ocurren después. Esa pregunta requiere otro lenguaje. Nadie terminó de inventarlo.

. . .

¿La bibliotecología va a seguir organizando lo que otros deciden, o está dispuesta a disputar quién decide qué conocimiento cuenta?

Referencias

- Argentina. (2013). Ley 26.899. Repositorios digitales institucionales de acceso abierto. Honorable Congreso de la Nación Argentina. <https://www.argentina.gob.ar/normativa/nacional/ley-26899-223459>
- Becerra, M., y Mastrini, G. (2009). *Los dueños de la palabra: acceso, estructura y concentración de los medios en la América latina del siglo XXI*. Prometeo.
- Beigel, F. (2017). Científicos periféricos, entre Ariel y Calibán. Saberes institucionales y circuitos de consagración en Argentina. Las publicaciones de los investigadores del CONICET. *DADOS, Revista de Ciências Sociais*, 60(3), 825-865. <https://doi.org/10.1590/001152582017136>
- Bender, E. M., Gebru, T., McMillan-Major, A., y Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? En *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623). ACM. <https://doi.org/10.1145/3442188.3445922>
- Busaniche, B. (Comp.). (2010). *Argentina copyleft. La crisis del modelo de derecho de autor y las prácticas para democratizar la cultura*. Fundación Vía Libre / Fundación Heinrich Böll. <http://sedici.unlp.edu.ar/handle/10915/15906>
- D'Amico, P. B., y Vercelli, A. (2024). Libraries, Access to Information and Artificial Intelligences in Argentina and Latin American and the Caribbean. *The International Information & Library Review*. <https://doi.org/10.1080/10572317.2024.2413775>
- Dodge, J., Sap, M., Marasovic, A., Agnew, W., Ilharco, G., Groeneveld, D., Mitchell, M., y Gardner, M. (2021). Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus. En *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 1286-1305). <https://doi.org/10.18653/v1/2021.emnlp-main.98>

- Kreimer, P. (2010). *Ciencia y periferia. Nacimiento, muerte y resurrección de la biología molecular en la Argentina. Aspectos sociales, políticos y cognitivos*. Eudeba.
- Kulesz, O. (2018). Culture, platforms and machines: the impact of artificial intelligence on the diversity of cultural expressions. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000380584>
- Lander, E. (Comp.). (2000). *La colonialidad del saber: eurocentrismo y ciencias sociales. Perspectivas latinoamericanas*. CLACSO.
- Mignolo, W. D. (2010). *Desobediencia epistémica: retórica de la modernidad, lógica de la colonialidad y gramática de la descolonialidad*. Ediciones del Signo.
- Quijano, A. (2000). Colonialidad del poder, eurocentrismo y América Latina. En E. Lander (Comp.), *La colonialidad del saber: eurocentrismo y ciencias sociales* (pp. 201-246). CLACSO.
- Ricaurte, P. (2022). Ethics for the Majority World: AI and the Question of Violence at Scale. *Media, Culture and Society*, 44(4), 726-745. <https://doi.org/10.1177/01634437221099612>
- Ricaurte, P., Gómez-Cruz, E., y Siles, I. (2024). Algorithmic governmentality in Latin America: Sociotechnical imaginaries, neocolonial soft power, and authoritarianism. *Big Data & Society*, 11(1). <https://doi.org/10.1177/20539517241229697>
- UNESCO. (2021). Recomendación sobre la ética de la inteligencia artificial. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa
- Vercelli, A. (2009). Repensando los bienes intelectuales comunes. Análisis sociotécnico sobre el proceso de co-construcción entre las regulaciones de derecho de autor y derecho de copia y las tecnologías digitales para su gestión [Tesis doctoral]. Universidad Nacional de Quilmes. <https://ridaa.unq.edu.ar/handle/20.500.11807/131>
- Vercelli, A. (2024). Regulaciones e inteligencias artificiales en Argentina. *InMediaciones de la Comunicación*, 19(1), 105-135. <https://doi.org/10.18861/ic.2024.19.1.3549>