Chapter 12 / Capítulo 12

Applied bibliometrics. From data to publication (Spanish Edition)
ISBN: 978-9915-9680-6-3
DOI: 10.62486/978-9915-9680-6-3.ch12

©2025 The authors. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY) 4.0 License.

EMERGING TRENDS / TENDENCIAS EMERGENTES

El panorama bibliométrico contemporáneo experimenta una transformación radical impulsada por la convergencia de tecnologías disruptivas que expanden sustancialmente las fronteras de lo analizable e interpretable. Estas innovaciones tecnológicas no solo optimizan metodologías establecidas, sino que fundamentalmente redefinen las preguntas de investigación que pueden formularse y responderse a través del análisis cuantitativo de la literatura científica. Desde la inteligencia artificial generativa hasta las tecnologías de registro distribuido, las tendencias emergentes prometen superar limitaciones históricas del campo mientras introducen nuevos desafíos éticos y metodológicos que la comunidad bibliométrica debe abordar con rigor y anticipación. Este capítulo examina críticamente estas tendencias transformadoras, evaluando tanto su potencial analítico como sus implicaciones para el futuro de la evaluación científica.

12.1. IA generativa: GPT en minería de textos científicos

Los modelos de lenguaje de gran escala como GPT representan una innovación disruptiva en el análisis bibliométrico al permitir la comprensión contextual de textos científicos a escala masiva. A diferencia de los métodos tradicionales de minería de texto basados en frecuencia de términos o análisis de co-ocurrencia, estos modelos capturan relaciones semánticas complejas y matices interpretativos que anteriormente requerían intervención humana especializada. Esta capacidad transforma sustancialmente el análisis de contenido bibliométrico, permitiendo desde la clasificación automática de artículos según dimensiones epistemológicas complejas hasta la identificación de patrones argumentativos y retóricos que trascienden la mera presencia de términos específicos. La aplicación de estos modelos a corpus bibliográficos completos abre posibilidades analíticas sin precedentes para comprender la evolución discursiva de los campos científicos.

La implementación de IA generativa en bibliometría permite la extracción automatizada de dimensiones analíticas tradicionalmente inaccesibles a través de métodos cuantitativos convencionales. Los modelos pueden identificar automáticamente contribuciones teóricas, innovaciones metodológicas y hallazgos empíricos clave en publicaciones científicas, clasificando artículos según taxonomías complejas sin requerir diccionarios terminológicos predefinidos. Esta flexibilidad analítica resulta particularmente valiosa para estudiar campos interdisciplinarios donde los vocabularios especializados evolucionan rápidamente y donde las contribuciones intelectuales trascienden las categorizaciones disciplinarias establecidas. La capacidad de estos sistemas para generar resúmenes analíticos sintetizando hallazgos de múltiples publicaciones amplifica aún más su utilidad para revisiones bibliométricas de grandes volúmenes de literatura.

El análisis de tendencias temáticas mediante IA generativa supera significativamente las limitaciones de los métodos basados en palabras clave al capturar evoluciones conceptuales y semánticas más allá de la mera frecuencia terminológica. Los modelos pueden rastrear cómo ciertos conceptos adquieren significados diferentes en distintos contextos temporales o disciplinarios, identificando procesos de recontextualización teórica o apropiación conceptual entre campos. Esta capacidad para mapear la deriva semántica de ideas científicas proporciona perspectivas únicas sobre la dinámica intelectual de los campos de investigación, revelando cómo los marcos conceptuales se transforman a través de su circulación al lado de diferentes comunidades epistémicas. Este análisis diacrónico de la evolución semántica constituye una contribución metodológica fundamental de la IA generativa a la bibliometría contemporánea.

La generación automática de hipótesis de investigación representa otra aplicación transformadora donde los sistemas de IA analizan patrones en la literatura científica para

identificar vacíos de conocimiento y oportunidades de investigación promisorias. Al procesar corpus documentales completos, estos sistemas pueden detectar conexiones no evidentes entre hallazgos aparentemente no relacionados, sugerir combinaciones innovadoras de enfoques metodológicos o identificar problemas de investigación infraexplorados con alto potencial de impacto. Esta capacidad de síntesis creativa al lado de dominios de conocimiento tradicionalmente separados posiciona a la IA generativa como herramienta valiosa para la planificación estratégica de la investigación y la identificación de fronteras científicas emergentes.

La evaluación de la calidad argumentativa y solidez metodológica mediante IA generativa introduce dimensiones cualitativas previamente inaccesibles a escala bibliométrica. Los modelos pueden analizar sistemáticamente la estructura argumental de las publicaciones, evaluar la adecuación entre preguntas de investigación y métodos empleados, e identificar limitaciones metodológicas recurrentes en literaturas específicas. Esta capacidad para evaluar dimensiones de calidad más allá de las citas recibidas ofrece alternativas prometedoras para sistemas de evaluación científica que trasciendan las métricas de impacto tradicionales, aunque introduce importantes desafíos de validación y transparencia algorítmica.

La implementación de IA generativa en bibliometría enfrenta desafíos significativos relacionados con la transparencia metodológica, los sesgos en los datos de entrenamiento y la reproducibilidad de los análisis. Los modelos de lenguaje exhiben tendencias a generar alucinaciones factuales, reproducir sesgos presentes en sus datos de entrenamiento y mostrar inconsistencia en respuestas al otro lado de diferentes solicitudes. Estos desafíos requieren el desarrollo de protocolos de validación rigurosos, la documentación exhaustiva de parámetros y *prompts* utilizados, y la implementación de mecanismos de explicabilidad que permitan comprender los fundamentos de las clasificaciones y análisis generados automáticamente.

La integración de IA generativa con métodos bibliométricos tradicionales establece un paradigma analítico híbrido que combina las fortalezas de ambos enfoques. Las técnicas establecidas de análisis de citas y redes proporcionan validación cuantitativa para las perspectivas cualitativas generados por IA, mientras el análisis semántico avanzado contextualiza y enriquece la interpretación de patrones estructurales identificados a través de métodos tradicionales. Esta integración metodológica produce análisis bibliométricos más robustos y matizados que aprovechan lo mejor de ambos paradigmas analíticos, estableciendo nuevos estándares para la comprehensión integral de la dinámica científica.

El futuro desarrollo de IA generativa aplicada a bibliometría probablemente se orientará hacia modelos especializados en dominios científicos específicos, entrenados en corpus disciplinarios comprehensivos y capaces de comprender convenciones retóricas y epistemológicas particulares de cada campo. La evolución hacia sistemas multimodales que integren análisis de texto, imágenes, datos y código en publicaciones científicas ampliará aún más las fronteras del análisis bibliométrico, permitiendo una comprensión más integral de la comunicación científica contemporánea. Esta especialización progresiva condiciona limitaciones actuales de los modelos generales mientras maximiza su utilidad para aplicaciones bibliométricas específicas.

12.2. *Blockchain* para métricas descentralizadas

La tecnología *blockchain* surge como una solución transformadora para abordar desafíos fundamentales de transparencia, confianza y descentralización en los sistemas de métricas académicas. A diferencia de las bases de datos centralizadas tradicionales, los registros distribuidos permiten crear sistemas de métricas incorruptibles donde cada transacción académica, citas, revisiones, contribuciones, queda registrada de manera inmutable y verificable

por cualquier participante de la red. Esta arquitectura descentralizada elimina puntos únicos de fallo y reduce la dependencia de intermediarios institucionales, estableciendo las bases para un ecosistema de métricas más resiliente y confiable. La aplicación de *blockchain* en bibliometría representa un cambio paradigmático hacia sistemas de evaluación científica más transparentes y resistentes a la manipulación.

La implementación de *blockchain* permite crear registros descentralizados de contribuciones académicas que capturan la totalidad del proceso investigador, no solo sus productos finales. Cada contribución, desde la conceptualización inicial y diseño metodológico hasta el análisis de datos y redacción, puede registrarse con marca temporal de manera inalterable en la cadena de bloques, creando trazas auditables de autoría y contribución intelectual. Este enfoque resuelve problemas crónicos de atribución de autoría y reconocimiento de contribuciones no tradicionales, particularmente valioso en investigación colaborativa donde múltiples investigadores contribuyen en diferentes capacidades a lo largo del ciclo de investigación. La granularidad en el registro de contribuciones permite métricas más justas y precisas que reflejan la naturaleza compleja y colaborativa de la ciencia contemporánea.

Los sistemas de métricas basados en *blockchain* introducen mecanismos innovadores para verificación de citas y prevención de manipulación métrica. Cada cita puede registrarse como una transacción en la cadena, creando un historial transparente e inmutable de relaciones de influencia intelectual. Este registro distribuido permite detectar patrones de citación circular, redes de auto citas coordinadas y otras prácticas manipulativas que distorsionan las métricas tradicionales. La verificación descentralizada mediante consenso de red elimina la dependencia de bases de datos centralizadas cuyos algoritmos de procesamiento y limpieza frecuentemente operan como cajas negras inescrutables para los usuarios finales.

Los tokens académicos y sistemas de reputación descentralizados representan otra aplicación disruptiva de *blockchain* en bibliometría. Estos sistemas permiten a investigadores acumular reputación basada en contribuciones verificadas registradas en la cadena, creando capital académico portable e independiente de instituciones específicas. Los mecanismos de tokenización pueden alinear incentivos individuales con objetivos colectivos del sistema científico, recompensando comportamientos académicos valiosos como revisiones por pares rigurosas, compartir datos de investigación o replicar estudios previos. Este enfoque transforma fundamentalmente los sistemas de recompensa académica al hacer explícitos y cuantificables valores tradicionalmente implícitos en la cultura científica.

La gestión descentralizada de identidades académicas mediante *blockchain* resuelve problemas persistentes de desambiguación y portabilidad de perfiles investigador. Cada investigador puede mantener una identidad auto soberana registrada en la cadena que consolida sus contribuciones a través de diferentes instituciones, plataformas y períodos temporales. Esta identidad persistente e independiente de proveedores específicos facilita la movilidad académica y elimina la fragmentación métrica que ocurre cuando investigadores cambian de afiliación institucional. La integración con sistemas ORCID existentes y otros identificadores académicos crea un ecosistema de identidad más robusto y centrado en el usuario.

Los contratos inteligentes permiten automatizar procesos de evaluación científica basados en métricas verificadas en la cadena. Estos protocolos auto ejecutables pueden implementar criterios de evaluación transparentes y automatizar procesos como revisión por pares, asignación de fondos o promociones académicas basadas en métricas objetivas registradas en *blockchain*. La automatización mediante contratos inteligentes reduce la carga administrativa en evaluación

científica mientras aumenta la consistencia y transparencia de los procesos decisorios. Sin embargo, esta automatización requiere diseño cuidadoso para evitar codificar sesgos existentes o crear sistemas excesivamente rígidos que no capturen dimensiones cualitativas importantes de la excelencia científica.

La implementación de métricas descentralizadas enfrenta desafíos significativos de escalabilidad, adopción y gobernanza. El registro de cada transacción académica en la cadena genera volúmenes masivos de datos que deben procesarse eficientemente, mientras la adopción por comunidades académicas diversas requiere superar inercias institucionales y estandarizar protocolos a través de diferentes disciplinas y regiones. Los modelos de gobernanza para estos sistemas descentralizados deben balancear participación comunitaria con eficiencia decisional, evitando tanto la captura por intereses particulares como la parálisis por exceso de deliberación. Estos desafíos prácticos requieren atención cuidadosa durante el diseño e implementación de sistemas bibliométricos basados en *blockchain*.

La integración de *blockchain* con otras tecnologías emergentes como inteligencia artificial e internet de las cosas crea ecosistemas métricos aún más sofisticados y comprehensivos. Los sistemas de IA pueden analizar los datos inmutables registrados en la cadena para generar conocimientos más confiables, mientras dispositivos *Internet of Things* (IoT) pueden automatizar el registro de actividades investigadoras en laboratorios y entornos de campo. Esta convergencia tecnológica potencia las capacidades de cada tecnología individual mientras mitiga sus limitaciones específicas, creando infraestructuras métricas más robustas, transparentes y alineadas con las prácticas reales de investigación científica en diferentes dominios y metodologías.

12.3. Análisis de imágenes y datos multimodales

El análisis de imágenes y datos multimodales representa una frontera emergente en bibliometría que expande significativamente el alcance de lo cuantificable en la comunicación científica. Tradicionalmente, los estudios bibliométricos se han centrado predominantemente en el análisis de texto, dejando de lado el rico contenido visual que acompaña a las publicaciones académicas contemporáneas. Las imágenes, gráficos, diagramas y visualizaciones constituyen componentes esenciales del discurso científico moderno, transmitiendo información compleja que frecuentemente no puede expresarse adecuadamente mediante solo texto. La capacidad de analizar sistemáticamente estos elementos visuales abre nuevas dimensiones para comprender cómo se construye y comunica el conocimiento en diferentes disciplinas científicas.

Las técnicas de visión artificial y aprendizaje profundo permiten extraer información significativa de elementos visuales en publicaciones científicas a escala masiva. Los algoritmos pueden identificar y clasificar automáticamente diferentes tipos de imágenes, desde gráficos estadísticos y diagramas de flujo hasta microfotografías y visualizaciones tridimensionales, estableciendo patrones de uso según disciplinas, metodologías o épocas específicas. Esta capacidad analítica transforma las imágenes de meras ilustraciones en datos cuantificables que revelan preferencias representacionales, estándares de visualización y evoluciones en las prácticas de comunicación visual dentro de diferentes comunidades científicas. El análisis automatizado de estos elementos visuales complementa los enfoques textuales tradicionales, ofreciendo una perspectiva más comprehensiva de la retórica visual en la ciencia.

El análisis multimodal integra información proveniente de diferentes formatos, texto, imágenes, tablas, ecuaciones, para generar entendimientos más ricos y contextualizados del contenido científico. En lugar de analizar cada modalidad por separado, los enfoques

multimodales capturan las interrelaciones entre diferentes formas de representación del conocimiento, revelando cómo los investigadores combinan recursos semióticos diversos para construir argumentos científicos persuasivos. Esta integración permite, por ejemplo, analizar cómo las descripciones textuales se relacionan con las visualizaciones correspondientes, o cómo las ecuaciones matemáticas se complementan con explicaciones narrativas y representaciones gráficas. La bibliometría multimodal captura así la naturaleza esencialmente híbrida de la comunicación científica contemporánea.

La detección de tendencias metodológicas a través del análisis de imágenes constituye una aplicación particularmente valiosa para el mapeo científico. Ciertos tipos de imágenes funcionan como indicadores confiables de enfoques metodológicos específicos, las microfotografías electrónicas sugieren ciertas técnicas de laboratorio, mientras los diagramas de circuitos señalan aproximaciones experimentales particulares en ingeniería. El análisis diacrónico de la prevalencia de diferentes tipos de imágenes puede revelar la adopción, consolidación o declive de metodologías de investigación, proporcionando perspectivas sobre la evolución de las prácticas científicas que complementan los análisis basados en terminología textual. Esta capacidad para rastrear metodologías a través de sus huellas visuales representa una contribución única del análisis multimodal a la bibliometría.

La evaluación de la calidad y claridad visual en publicaciones científicas emerge como otra aplicación prometedora del análisis multimodal. Los algoritmos pueden evaluar aspectos como la resolución de imágenes, la adecuación de las escalas, la claridad de las etiquetas y la efectividad comunicativa de diferentes tipos de visualizaciones. Esta evaluación automatizada de la calidad visual permite estudios a escala de la evolución de los estándares de visualización científica, la identificación de mejores prácticas en comunicación visual y el análisis de cómo la calidad visual se correlaciona con el impacto científico medido a través de citas y otros indicadores. Esta dimensión cualitativa tradicionalmente subjetiva se transforma así en objeto de análisis cuantitativo sistemático.

El análisis de datos multimodales enfrenta desafíos metodológicos significativos relacionados con la estandarización, interpretación y validación de resultados. La enorme diversidad de formatos visuales en disciplinas científicas complica el desarrollo de taxonomías y algoritmos universalmente aplicables. Además, la interpretación de elementos visuales frecuentemente requiere conocimiento especializado del dominio, haciendo difícil la automatización completa del análisis sin perder matices disciplinares cruciales. La validación de los resultados del análisis multimodal requiere estrategias innovadoras que combinen evaluación automática con verificación por expertos humanos, estableciendo estándares de confiabilidad para esta nueva frontera bibliométrica.

La integración del análisis multimodal con otras tendencias emergentes como la IA generativa y *blockchain* crea sinergias particularmente poderosas. Los modelos de lenguaje multimodal pueden generar descripciones textuales de contenido visual o crear visualizaciones a partir de descripciones textuales, facilitando la accesibilidad y búsqueda de contenido científico multimodal. Simultáneamente, *blockchain* puede proporcionar la infraestructura para registrar y verificar la procedencia de imágenes y datos multimodales, direccionando preocupaciones crecientes sobre la manipulación de imágenes científicas. Estas integraciones tecnológicas amplifican el potencial transformador del análisis multimodal en bibliometría.

El futuro desarrollo del análisis bibliométrico multimodal probablemente evolucionará hacia sistemas capaces de capturar dimensiones cada vez más sofisticadas de la comunicación científica

integrada. La capacidad de analizar no solo imágenes estáticas sino también visualizaciones interactivas, videos científicos y representaciones inmersivas en realidad virtual expandirá aún más las fronteras de lo analizable.

12.4. Ética y transparencia en la era de la IA

La implementación de sistemas de inteligencia artificial en bibliometría plantea desafíos éticos fundamentales que requieren marcos de gobernanza robustos. La opacidad algorítmica característica de muchos modelos contemporáneos genera riesgos significativos de reproducir y amplificar sesgos históricos en la evaluación científica, perpetuando desigualdades estructurales. Esta falta de transparencia socava la legitimidad de los sistemas de evaluación bibliométrica basados en inteligencia artificial, exigiendo aproximaciones metodológicas que prioricen la explicabilidad y auditabilidad de los procesos decisorios para mantener la confianza de la comunidad académica en estas herramientas emergentes.

Garantizar la transparencia metodológica constituye un pilar esencial mediante la documentación exhaustiva de datos de entrenamiento, arquitecturas de modelos y procedimientos de validación empleados. Los metadatos deben incluir información detallada sobre la composición demográfica y geográfica de los conjuntos de datos utilizados, permitiendo identificar brechas representacionales que puedan distorsionar los resultados. Los investigadores tienen la responsabilidad ética de divulgar tanto capacidades como limitaciones conocidas de sus sistemas, evitando extrapolaciones indebidas más allá de las condiciones de desarrollo y validación originales, lo que facilita la evaluación crítica por pares.

La protección de datos personales emerge como consideración crítica cuando los sistemas procesan información sensible de investigadores, incluyendo trayectorias profesionales y redes de colaboración. Los marcos éticos deben asegurar el cumplimiento normativo, implementar principios de minimización en la recolección y establecer protocolos claros de consentimiento informado. Estas salvaguardas resultan particularmente importantes en contextos evaluativos donde los resultados pueden afectar oportunidades profesionales y asignación de recursos, requiriendo balances cuidadosos entre utilidad analítica y protección de la privacidad individual en el ecosistema investigador.

La rendición de cuentas por decisiones algorítmicas representa otro desafío central, exigiendo cadenas claras de responsabilidad cuando sistemas automatizados generan rankings o evalúan mérito científico. Deben establecerse mecanismos de apelación efectivos para casos donde investigadores o instituciones consideren evaluaciones injustas, complementados con comités de supervisión ética que incorporen diversidad disciplinaria y geográfica. Esta estructura de gobernanza distribuida permite identificar riesgos operativos y establecer protocolos para el uso responsable que mantengan la confianza institucional en estos sistemas emergentes de evaluación bibliométrica.

Mitigar proactivamente los sesgos algorítmicos demanda estrategias sistemáticas que incluyan desarrollo de técnicas especializadas, auditorías regulares de equidad y diversificación de equipos de desarrollo. La evaluación continua del impacto diferencial entre grupos demográficos, instituciones y regiones geográficas permite identificar y corregir disparidades emergentes antes de su consolidación como injusticias estructurales. Este enfoque preventivo resulta esencial para construir sistemas bibliométricos que reflejen los valores de equidad e inclusión propios de la empresa científica contemporánea en su diversidad epistemológica y metodológica.

La sostenibilidad ecológica completa el panorama ético, considerando los requerimientos computacionales masivos que generan huellas de carbono significativas durante el entrenamiento y despliegue de modelos. Los desarrolladores deben priorizar eficiencia energética, seleccionar arquitecturas balanceadas y transparentar costos ambientales, alineando la innovación tecnológica con los imperativos de responsabilidad ambiental que caracterizan la ciencia contemporánea. Esta conciencia ecológica representa una dimensión esencial de la ética aplicada al desarrollo de herramientas bibliométricas con inteligencia artificial en el contexto actual de crisis climática global.

La gobernanza participativa involucra múltiples actores-investigadores, gestores, representantes institucionales- en el diseño y supervisión de sistemas, asegurando que diversas perspectivas informen su desarrollo. Establecer procesos deliberativos abiertos sobre valores incorporados, compensaciones entre objetivos y estándares de rendición de cuentas construye consensos sociales alrededor del uso apropiado de estas tecnologías. Esta inclusividad mitiga riesgos de imposición tecnocrática y garantiza que los sistemas reflejen los valores plurales de la comunidad científica global en su conjunto diverso y multicultural.

La formación en ética digital desarrolla usuarios críticos capaces de interpretar resultados algorítmicos mediante programas que enfaticen el juicio humano como complemento esencial de evaluaciones automatizadas. Cultivar escepticismo informado hacia resultados contraintuitivos o potencialmente sesgados previene aplicaciones acríticas, mientras se fortalecen capacidades comunitarias para aprovechar beneficios y mitigar riesgos. Esta alfabetización constituye una inversión necesaria en el capital humano que utilizará estas herramientas, asegurando su implementación responsable en los diversos contextos institucionales y disciplinares donde se aplicarán.

Recapitulando

- La ciencia abierta promueve transparencia y reproducibilidad.
- Las plataformas OpenAlex y Crossref democratizan el acceso a metadatos.
- La inteligencia artificial facilita el análisis y predicción de tendencias.
- La altmetría amplía la medición del impacto más allá de las citas.
- El uso de paneles interactivos mejora la visualización dinámica.
- La transparencia algorítmica es fundamental en bibliometría basada en IA.
- Big data permite analizar producción científica en tiempo real.
- Los indicadores de equidad evalúan diversidad de género, idioma y geografía.
- Integrar patentes, publicaciones y financiamiento mejora evaluación de innovación.
- Herramientas reproducibles como R y Python facilitan replicación.
- Riesgos: manipulación algorítmica y *gaming* de métricas.
- Manifiestos como DORA y Leiden impulsan evaluación responsable.
- Los repositorios abiertos fortalecen la ciencia colaborativa.
- La interoperabilidad de datos (DOI, ORCID) mejora trazabilidad.
- Las métricas alternativas reflejan impacto social y mediático.
- Las paneles web permiten seguimiento continuo de indicadores.
- Los informes métricos evolucionan hacia visualizaciones narrativas.
- La ética de IA se convierte en prioridad en investigación.
- La bibliometría predictiva combina algoritmos y vigilancia de tendencias.
- La capacitación en herramientas de análisis se vuelve esencial en el futuro.

Preguntas de autoevaluación

1. ¿Qué cambios aporta la IA al análisis bibliométrico?

2. ¿Qué diferencia existe entre altmetría y citaciones tradicionales?
3. ¿Qué fuentes abiertas son relevantes para nuevos estudios?
4. ¿Cuáles son los principales riesgos éticos de la IA en evaluación científica?
5. ¿Qué promueve la Declaración DORA?
6. ¿Por qué es importante la interoperabilidad de datos?
7. ¿Qué ventajas ofrecen los paneles interactivos?
8. ¿Qué implica el concepto de evaluación responsable?
9. ¿Qué herramientas son claves para análisis reproducibles?
10. ¿Cómo puede la bibliometría predecir tendencias emergentes?

BIBLIOGRAFÍA

1. Sugimoto CR, Larivière V. Measuring research: what everyone needs to know. Oxford: Oxford University Press; 2018. ISBN 9780190640125

2. Thelwall M. Web indicators for research evaluation: a practical guide. San Rafael (CA): Morgan & Claypool; 2016. https://doi.org/10.2200/S00733ED1V01Y201602ICR048

3. Gingras Y. Bibliometrics and research evaluation: uses and abuses. Cambridge (MA): MIT Press; 2016. ISBN 9780262337663

4. Wilsdon J, et al. The metric tide: independent review of the role of metrics in research assessment and management. London: HEFCE; 2015. ISBN 9780992492124