

Análisis bibliométrico de la producción científica en Inteligencia Artificial en Argentina

Braian Tadei

braiantadei@gmail.com

Mesa: Tesinas y prácticas de licenciatura en curso.

Directoras: Claudia M. González cgonzalez@fahce.unlp.edu.ar

Sandra Miguel smiguel@fahce.unlp.edu.ar

Palabras clave: BIBLIOMETRÍA; CIENCIOMETRÍA; INTELIGENCIA ARTIFICIAL; ARGENTINA; PRODUCCIÓN CIENTÍFICA.

Resúmenes ampliados

Definición de la problemática

La Inteligencia Artificial (AI) es una disciplina de las Ciencias de la Computación de relevancia cada vez mayor. Sus aplicaciones se están expandiendo hacia los más diversos aspectos de la vida social, convirtiéndola en una “tecnología de propósito general” (Albrieu et al., 2018). En la actualidad, se considera que la AI podría llevar a la automatización, ya no de procesos simples o repetitivos, sino de tareas complejas y no rutinarias (Aghion, Jones y Jones, 2017). En Argentina, ha sido tomada como área estratégica en el desenvolvimiento de la política científica, existen centros de investigación especializados dependientes de CONICET y de Universidades Nacionales, y se ha dado impulso a sucesivos programas estatales de incentivo al desarrollo de AI, desde 2019 a la fecha (Argentina, 2019). No existe hasta el momento, sin embargo, un estudio pormenorizado y actualizado del volumen y las características de la producción científica de investigadores/as argentinos/as en esta área del saber.

La presente tesina de Licenciatura, de elaboración en curso, se plantea indagar en detalle qué ocurre con la investigación científica en el campo de la AI, relevando para ello los documentos incluidos en la base de datos Scopus. ¿Cuál es el volumen de esta producción, y cuánto representa en relación a la producción científica argentina en general? ¿Cómo ha sido el crecimiento de la producción en AI en años recientes? ¿Cuáles son los temas de investigación? ¿Cuáles son las personas y las instituciones más destacadas? ¿Qué medios prefieren para publicar quienes investigan en el área?

¿Cuáles son las redes de colaboración, entre personas, instituciones o países? Por último, ¿cuál es el impacto de lo publicado?

Objetivos

El objetivo general del trabajo es revelar información de carácter bibliométrico que contribuya al conocimiento de las actividades de investigación que se llevan adelante en Argentina sobre Inteligencia Artificial (AI).

Los objetivos específicos incluyen: caracterizar la producción científica argentina en AI incluida en Scopus utilizando indicadores de crecimiento e impacto; analizar las redes de colaboración nacional e internacional; identificar las áreas temáticas que componen el núcleo de la AI en Argentina y las disciplinas con las que se relaciona; identificar a los/las investigadores/as, e instituciones de referencia para el campo en nuestro país; identificar las principales fuentes utilizadas para publicar.

Enfoque metodológico

La metodología de trabajo se plantea en dos etapas. En primer lugar, la delimitación del campo de la Inteligencia Artificial, necesaria para construir una estrategia de búsqueda que permita relevar con exhaustividad y precisión la producción científica objetivo (Muñoz-Écija, T., Vargas Quesada, B. y Chinchilla Rodríguez, Z., 2019). El propósito de esta fase es obtener un conjunto de términos que sirvan para la recolección de documentos en las bases de datos. En segundo lugar, luego de recuperados los registros bibliográficos, es la etapa del procesamiento y análisis métrico. Para esta segunda fase, se utilizará la herramienta Bibliometrix, un paquete desarrollado en el lenguaje de programación R, destinada a facilitar la obtención, análisis y presentación de diversos indicadores bibliométricos (Aria, Cuccurullo, 2017). Para este trabajo se optó relevar los registros bibliográficos incluidos en la base de datos Scopus, por ser una de las BBDD bibliográficas de mayor tamaño, con alcance mundial (Vuotto, Di Césare, Pallotta, 2020), y por ser accesible a través del convenio de la Biblioteca Electrónica de Ciencia y Tecnología del Ministerio de Ciencia, Tecnología e Innovación de la República Argentina (MinCYT).

Delimitación de campo y obtención de keywords

Los esfuerzos por modelar la delimitación del campo de la AI son desarrollados en un informe publicado por Elsevier en 2018, titulado “Artificial Intelligence: How knowledge is created, transferred, and used”, donde se analizan cientos de miles de documentos que hacen referencia a la AI desde cuatro enfoques: el ámbito científico, el educativo, el de la opinión pública, y el de la industria. De ellos se extraen las principales palabras clave, combinando sucesivamente una serie de técnicas: text mining sobre los títulos y resúmenes; uso de tesauros y procesadores de lenguaje natural; y el análisis caso por caso con consulta a expertos para determinar la pertinencia de las expresiones obtenidas. El estudio nos proporciona una lista de 232 palabras clave relevantes, disponibles en el dashboard del centro de recursos sobre AI de Elsevier (“AI Resource Center”, disponible en: <https://www.elsevier.com/connect/resource-center/artificial-intelligence>). En el marco de este trabajo, la lista se utilizará como principal insumo para la elaboración de la estrategia de búsqueda en Scopus, previa depuración y organización.

Al momento de crear el corpus de palabras clave para la búsqueda y relevamiento de documentos, se tuvieron presente la relevancia y la especificidad de los términos para hacer una revisión de la lista de keywords obtenida de Elsevier. Este trabajo fue realizado siguiendo dos métodos complementarios: en primer lugar, una revisión caso por caso de las 232 keywords, detectando las que corresponden a términos específicos del campo, y analizando los no específicos para determinar si fueran o no relevantes, apoyando el análisis en materiales bibliográficos de referencia junto con recursos web especializados en AI, su enseñanza y su divulgación. En segundo lugar, una revisión automatizada utilizando Chat GPT-3, guiando la interfaz mediante preguntas, se le pidió al procesador que clasifique los términos en tres categorías: a) Términos relevantes y específicos; b) términos relevantes, pero no específicos; c) términos que pertenecen a otros campos, pero pueden ser relevantes en una mirada amplia sobre IA.

La intención con esta doble operación es comparar los resultados y obtener una doble verificación. De los 232 términos revisados, en 84 (es decir, cerca del 32%) hubo una diferencia de caracterización entre el análisis caso por caso y el análisis de GPT. Las potencialidades de esta tecnología para la asistencia al trabajo intelectual, y específicamente para la investigación científica entra aún en la categoría de lo indefinido, pues se trata de un campo en pleno desarrollo y podemos atestiguar sus transformaciones, avances y limitaciones en tiempo real (Lund y Wang, 2023).

Estrategia de búsqueda

Para la estrategia de búsqueda, se encadenaron las keywords mediante conectores y operadores booleanos, se descartaron términos redundantes, se agregaron otros obtenidos de la literatura de referencia, y los términos detectados como no específicos fueron combinados con otras keywords para asegurar su especificidad respecto del campo. La búsqueda también fue limitada para que incluya sólo artículos, conferencias, libros y capítulos de libros, con al menos un autor con Argentina como país de afiliación, y con año máximo de publicación 2022 (último año con información completa).

La importancia del análisis terminológico y el énfasis en la elaboración de la fórmula de búsqueda se ve en los resultados de búsquedas comparadas. La estrategia de búsqueda avanzada arroja 4953 documentos, lo cual se juzga un volumen apto para realizar análisis bibliométricos. Una búsqueda únicamente a través del término “artificial intelligence” (manteniendo los filtros por tipo de documento y afiliación en Argentina), devuelve sólo 879 resultados: es decir, un 82% de los documentos recuperados mediante la búsqueda avanzada quedarían por fuera del análisis.

Análisis y primeros resultados

Los resultados permiten caracterizar que la investigación de científicos/as argentinos/as en este campo tiene una trayectoria de casi cuatro décadas. Unos pocos documentos publicados al año durante la década de 1980 marcan una primera etapa de aparición incipiente. Comienza un crecimiento sostenido a mediados de los 90, y nuevos saltos cuantitativos rondando los años 2000, 2010 y 2020.

Del total de documentos recuperados, 3078 son artículos de revistas científicas, mientras que 1822 son documentos de conferencias, 99 son capítulos de libros y sólo 6 son libros. La gran mayoría de ellos fueron publicados en inglés.

Medido a través de la citación, el impacto de esta producción científica también crece, configurando etapas cronológicas similares. El promedio de citas recibidas anualmente por cada documento pasa de 1 en las décadas del 80 y 90, 2 entre 2001 y 2012, y de 3 de 2013 en adelante.

Los resultados preliminares también permiten ver una evolución temática

marcada. Mientras las keywords relacionadas con las ciencias exactas, la estadística, las matemáticas y la computación tienen preponderancia en la primera etapa, a partir de 2010 adquieren preponderancia las palabras clave relacionadas con desarrollos y aplicaciones específicas de la IA, como machine learning y deep learning, procesamiento de lenguaje natural, industria 4.0 o computer vision.

Referencias bibliográficas.

1. Albrieu, R., Rapetti, M., Brest López, C., Larroulet, P. y Sorrentino, A. (2018). *Inteligencia artificial y crecimiento económico. Oportunidades y desafíos para Argentina*. CIPPEC. <https://www.cippec.org/wp-content/uploads/2018/11/ADE-ARG-vf.pdf>
2. Aghion, P., Jones, B. F., y Jones, C. I. (2017). *Artificial Intelligence and Economic Growth*. National Bureau of Economic Research. National Bureau of Economic Research. <https://www.nber.org/papers/w23928>
3. Argentina (2019). *ArgenIA. Plan Nacional de Inteligencia Artificial*. <https://ia-latam.com/wp-content/uploads/2020/09/Plan-Nacional-de-Inteligencia-Artificial.pdf>
4. Aria, M. y Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975. DOI: <https://doi.org/10.1016/j.joi.2017.08.007>
5. Elsevier (2018). *Artificial Intelligence: How knowledge is created, transferred, and used. Trends in China, Europe, and the United States*. Elsevier. <https://www.elsevier.com/research-intelligence/resource-library/ai-report>
6. Elsevier Connect (2020). *AI Resource Center: Free access to research and expert commentary on artificial intelligence (AI) and related discussions*. <https://www.elsevier.com/connect/resource-center/artificial-intelligence>
7. Lund, B. D., y Wang, T. (2023). Chatting about ChatGPT: How may AI and GPT impact academia and libraries? [Manuscrito presentado para su publicación]. *ResearchGate*. DOI: <https://doi.org/10.1108/LHTN-01-2023-0009>
8. Muñoz-Écija, T., Vargas Quesada, B. y Chinchilla Rodríguez, Z. (2019). Coping with methods for delineating emerging fields: Nanoscience and nanotechnology as a case study. *Journal of Informetrics*, 13(4). DOI: <https://doi.org/10.1016/j.joi.2019.100976>
9. Vuotto, A., Di Césare, V. y Pallotta, N. (2020). Fortalezas y debilidades de las

principales bases de datos de información científica desde una perspectiva bibliométrica. *Palabra Clave*, 10(1). DOI: <https://doi.org/10.24215/18539912e101>