

La producción científica mexicana en Inteligencia Artificial: un análisis bibliométrico

Gabriela Elisa Sued*

Artículo recibido:
10 de febrero de 2024
Artículo aceptado:
2 de mayo de 2024

Artículo de investigación

RESUMEN

Este artículo releva el estado actual de la producción científica de la IA en México con técnicas bibliométricas. Considera la especialización de la IA en seis subcampos. Como metodología, emplea los metadatos de 13 265 publicaciones, recolectados del catálogo bibliográfico OpenAlex y realiza un análisis cuantitativo de productividad con base en métricas de publicaciones, autores, citas y colaboraciones internacionales, e identifica sus principales temas de investigación y su desarrollo. Los resultados muestran una estructura científica local extensa con importantes colaboraciones internacionales. Se identifican tanto subcampos maduros, desarrollados desde hace tres décadas, que abarcan la robótica y

* Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, México gabriela.sued@iimas.unam.mx

las redes neuronales; como subcampos emergentes, desarrollados en los últimos cinco años, que comprenden el aprendizaje automático, el procesamiento del lenguaje natural y la visión por computadora. El artículo distingue aplicaciones recientes en las áreas de salud, medio ambiente, finanzas, procesamiento del lenguaje natural y acústica.

Palabras clave: Inteligencia Artificial; Bibliometría; Producción Científica; México

Mexican Scientific Production on Artificial Intelligence: A Bibliometric Analysis

Gabriela Elisa Sued

ABSTRACT

This article surveys the current state of AI scientific production in Mexico using bibliometric techniques. It examines AI specialization across six subfields. As a methodology, it utilizes metadata from 13 265 publications—collected from the OpenAlex catalogue—and conducts a quantitative productivity analysis based on metrics of publications, authors, citations, and international collaborations, identifying key research themes and their development. Findings reveal a broad local scientific structure with significant international collaborations. Both mature subfields, developed over three decades, including robotics and neural networks, and emerging subfields, developed in the last five years, encompassing machine learning, natural language processing, and computer vision, were identified. The article highlights recent applications in the health, environment, finance, natural language processing, and acoustics fields.

Keywords: Artificial Intelligence; Bibliometrics; Scientific Production; Mexico

INTRODUCCIÓN

La Inteligencia Artificial (IA) es un sistema sociotécnico complejo en el que intervienen actores heterogéneos como centros de investigación, empresas, usuarios y gobiernos. Se define como la habilidad de un sistema para aprender de

datos externos correctamente para interpretarlos y lograr objetivos específicos a través de una adaptación flexible (Kaplan y Haenlein, 2019: 3). Se divide en subcampos con objetos de estudio, métodos y objetivos diferenciados: el aprendizaje automático, la robótica, la visión por computadora, las redes neuronales, el procesamiento del lenguaje natural y el aprendizaje profundo son los principales. Cada subcampo da lugar a aplicaciones que transforman la vida cotidiana y social. Por ejemplo, la robótica incide no solo en la fabricación de automóviles, sino también en su conducción; las aplicaciones de procesamiento de lenguaje humano, junto con los agentes inteligentes, asisten a usuarios de Internet para ejecutar tareas que implican la elaboración de texto e imágenes; y el aprendizaje automático ha comenzado a usarse para la detección y el diagnóstico temprano de enfermedades, la administración de la energía, el reconocimiento de voz e imagen y en la clasificación de información (Girasa, 2020: 33-39).

El desarrollo de IA es tanto transnacional, a cargo de grandes compañías tecnológicas, como local. En América Latina la labor de los centros de investigación establecidos en cada país es relevante. Varios organismos internacionales observan el desarrollo y la aplicación de IA en toda la región, en ocasiones como un factor de inclusión e innovación para el bien común, en otras a través de una preocupación por la pérdida de puestos en diferentes sectores laborales, lo que podría balancearse con nuevas oportunidades de creación de empleo (Gómez-Mont *et al.*, 2020: 13).

La historia de la IA en México comenzó con la instalación de la primera computadora en la Facultad de Ciencias de la Universidad Nacional Autónoma de México (UNAM) en 1958. En 1986 se creó la Sociedad Mexicana de Inteligencia Artificial, y en 1994, el primer Centro de Investigación en IA, así como el primer posgrado en la Universidad Veracruzana (UV). En las dos últimas décadas tomó impulso con la creación de la primera carrera de grado en 2007, del primer doctorado en 2012 y con la organización de jornadas y congresos (Vergara-Villegas *et al.*, 2021: 2-3). No existe, sin embargo, una revisión sistemática de la investigación científica desarrollada en México. Esta información es clave tanto para planificar agendas de investigación, como para impulsar el desarrollo económico, promover vocaciones e implementar lineamientos éticos y regulatorios.

La IA es una tecnología emergente que altera significativamente sistemas sociales y productivos, e interactúa con otras disciplinas. Por su dinamismo, las tecnologías emergentes necesitan un análisis paralelo a su desarrollo y un repertorio de metodologías innovadoras. La bibliometría, por ejemplo, ha contribuido en la creación de indicadores para evaluar tecnologías emergentes como la bio y la nanotecnología (Amaro-Rosales y Robles-Belmont, 2020: 17).

Este artículo se propone relevar el estado actual de la producción científica de la IA en México con el objeto de contribuir a la identificación de capacidades

científicas y a la elaboración de indicadores para la evaluación de tecnologías emergentes, a través de un análisis bibliométrico (Donthu *et al.*, 2021: 285). Se pregunta acerca de las tendencias de crecimiento de los subcampos que componen la investigación mexicana en IA; de la estructura intelectual emergente del análisis de las publicaciones, citaciones, colaboraciones internacionales, experiencia y género de sus autores; e indaga en sus principales temas de investigación, su progreso y las áreas de aplicación que han surgido en los últimos años.

Aunque la producción de IA en México no ha sido estudiada desde un enfoque bibliométrico, hay investigaciones relevantes. Lancho-Barrantes y Cantú-Ortiz (2019: 501-502) revisaron la producción científica de México entre 2007 y 2017 a través de 177 574 publicaciones indexadas en Scopus y clasificadas por disciplinas. En este artículo, la cantidad de trabajos en ciencias de la computación es mucho menor respecto de las disciplinas que ocupan los primeros lugares (506). Sin embargo, los autores destacan a Oscar Castillo, Patricia Melin, investigadores en redes neuronales artificiales, y Leonid Fridman, del área de robótica, como los investigadores más citados en toda la base de datos (509). Arencibia-Jorge *et al.* (2023: 10) usaron técnicas bibliométricas para analizar la producción científica mexicana en varias disciplinas, observando en la IA patrones similares a los globales.

Ruiz León (2018: 17) se enfocó en la producción del Departamento de Ciencias de la Computación en el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la UNAM, destacando que la categoría IA presenta la mayor cantidad de documentos publicados. Cornejo *et al.* (2023: 1435) revisaron el estado de la robótica y la automatización en México, resaltando su relevancia a nivel global y su vinculación con la industria manufacturera. López-Martínez y Sierra (2021: 189) realizaron un análisis bibliométrico de la investigación en procesamiento de lenguaje natural; examinaron 521 publicaciones entre 2004 y 2017 y presentaron un análisis de palabras concurrentes donde resaltaban las temáticas de “búsqueda de información”, “elaboración automática de sumarios” y “clasificación de textos”. El aprendizaje automático tenía una presencia marginal, ya que sus términos asociados –“detección de sentimientos”, “programación genética” y “algoritmos evolutivos”– exhibían una baja frecuencia; tampoco encontraron publicaciones que relacionaran el procesamiento de lenguaje natural con el aprendizaje profundo.

Según Loan, Bashir y Nasreen (2021: 29), en los últimos años se han publicado pocos análisis bibliométricos sobre IA, pese al constante crecimiento de trabajos en el área. Los autores condujeron un análisis bibliométrico de 1 109 artículos publicados en la revista *Applied Artificial Intelligence*, entre 1991 y 2019 (31). Encontraron que los cinco conceptos más frecuentes en las publicaciones eran “model”, “system”, “classification”, “optimization” y “algorithm”. La frecuencia de los conceptos asociados con “neural networks” estaba por debajo del décimo lugar (39).

El presente trabajo prosigue con la descripción de la metodología y continúa con los hallazgos, divididos en dos secciones: una sobre métricas de desempeño académico y otra sobre mapeo temático. La discusión expone los puntos clave del análisis y la conclusión los contrasta con la literatura previa, además de sugerir futuras líneas de investigación.

METODOLOGÍA

La bibliometría emplea los metadatos de las publicaciones académicas disponibles en grandes catálogos bibliográficos como Scopus, Web of Science u OpenAlex (Bredahl, 2022: 5). La base recolectada está compuesta por los metadatos bibliográficos de todos los trabajos que reportan por lo menos una institución mexicana en la filiación de los autores. Se utilizaron las seis claves de búsqueda en inglés caracterizadas en la *Tabla 1*. La búsqueda fue aplicada a títulos, resúmenes y conceptos; se restringió al área de conocimiento de ‘Ciencias de la computación’.

Clave de búsqueda	Traducción al español	Definición
Machine & Deep Learning	Aprendizaje automático (AA) y Aprendizaje profundo (AP)	Entrenamiento de dispositivos o software para la ejecución de una tarea. El AP es un subcampo dentro del AA que usa algoritmos conectados en capas, de manera análoga al cerebro humano.
Natural Language Processing	Procesamiento del Lenguaje Natural (PLN)	Subcampo enfocado en la relación entre el lenguaje humano-computadora y las formas en las que máquinas procesan grandes cantidades de datos lingüísticos.
Speech Processing	Procesamiento del habla (PH)	Conversión del habla humana a señales en formato digital.
Computer Vision	Visión por computadora (VC)	Registro, procesamiento, análisis y comprensión de imágenes en un rango amplio de actividades. Asociada a la robótica, los procesos industriales y el reconocimiento facial.
Artificial Neural Networks	Redes Neuronales Artificiales (RNA)	Diseño de sistemas computacionales contruidos con un número simple de procesadores altamente conectados, basados en redes neuronales biológicas.
Evolutionary Computing & Genetic Algorithms	Computación Evolutiva (CE) y Algoritmos Genéticos (AG)	Subcampo basado en los sistemas evolutivos de las especies para el diseño de soluciones computacionales para problemas complejos.

Robots & Robotics	Robots y Robótica (ROB)	Subcampo aplicado a la teoría, diseño y fabricación de robots.
Artificial Intelligence	Inteligencia Artificial (IA)	Se usó la categoría general para agrupar los trabajos que no entraban en los subcampos definidos.

Tabla 1. Definición de subcampos de la IA
Fuente: elaboración de la autora con base en Girasa (2020: 13-21)

Para definir el catálogo bibliográfico que sería usado se comparó la cantidad de trabajos disponibles por subcampos en las colecciones Scopus, Web of Science y OpenAlex (*Tabla 2*).

Clave de búsqueda	Scopus	Web of Science	OpenAlex
Machine Learning OR Deep Learning	1888	2349	2566
Natural Language Processing OR Speech Processing	631	330	511
Computer Vision OR Digital Image Processing	66	470	1012
Neural Networks OR Evolutionary Computing OR Evolutive Algorithms	3673	154	4955
Robots OR Robotics	449	4905	3713
Artificial Intelligence	2993	920	510
Total	9700	9128	13267

Tabla 2. Trabajos disponibles en diferentes catálogos bibliográficos
Fuente: elaboración de la autora

Entre los catálogos revisados, OpenAlex (OurResearch, 2022) es el único de acceso abierto y el que proporcionó la mayor cantidad de artículos. Este alberga más de 240 millones de trabajos académicos, con un aumento diario de más de cincuenta mil. Cada trabajo en OpenAlex es clasificado automáticamente con un índice de conceptos aplicados a títulos y resúmenes (Priem, Piwowar y Orr, 2022: 3).

Para recolectar datos, se accedió a la API de OpenAlex mediante la librería “openalexR” (Aria *et al.*, 2023) en *RStudio* (RStudio Team, 2020) durante septiembre de 2023. Se capturaron 13 265 metadatos de trabajos publicados desde su inicio hasta 2022, aunque los registros en la base de datos se vuelven consistentes y continuos a partir del 2000. Los trabajos que figuraban en varios subcampos fueron clasificados en aquel que hubiera obtenido la mayor puntuación de confianza adjudicada por OpenAlex al realizar la asignación automática de conceptos.

Para el procesamiento de datos se emplearon las librerías “tidytext” (Queiroz *et al.*, 2023), “ggplot2” (Wickham *et al.*, 2023c), “tidyr” (Wickham *et al.*, 2023a) y “dplyr” (Wickham *et al.*, 2023b). La distribución de autorías por género fue realizada con la API Genderized.io (Demografix ApS, s.f.), que asigna un género a los autores según su nombre de pila con una precisión de 90%. Para modelar las redes temáticas se emplearon Gephi (Bastian, Heymann y Jacomy, 2009: 361) y *VosViewer* (Eck y Waltman, 2019). Las bases de datos y el código utilizado fueron publicados en un repositorio de acceso abierto para su consulta y reúso (Sued, 2024a; 2024b).

RESULTADOS

Las subsecciones siguientes exponen los resultados del procesamiento de datos. La primera presenta un análisis de desempeño y la segunda, un mapeo temático (Donthu *et al.*, 2021: 288).

Análisis de desempeño

La *Figura 1* muestra las métricas de desempeño de publicaciones por subcampo desde el 2000. El número de publicaciones ha aumentado progresivamente, alcanzando su pico en 2022. Se observan patrones distintos en cada subcampo: ROB, RNA y CE lideran el número de publicaciones:

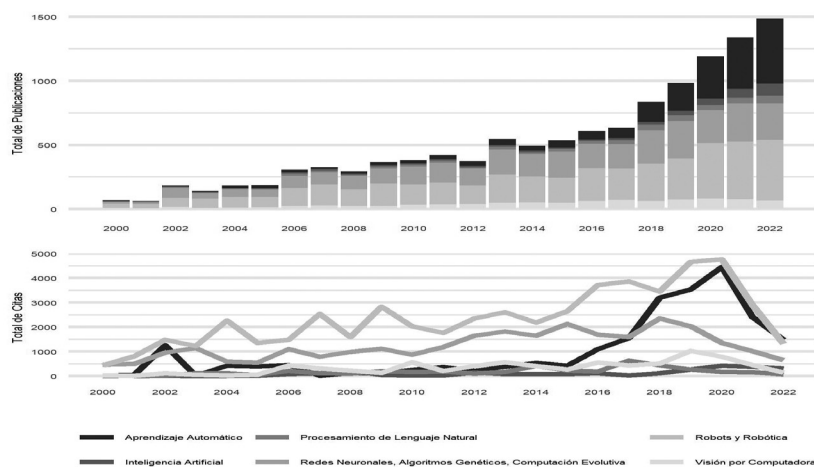


Figura 1. Trabajos publicados y citas por subcampo
Fuente: elaboración de la autora

El AA evidencia una expansión reciente y rápida. En los subcampos PLN y VC se producen relativamente pocas publicaciones, aunque aumentan progresivamente, sobre todo en VC. IA como categoría general registra menos publicaciones, lo que puede suponer una especialización significativa y áreas de trabajo bien definidas. Las citas también han crecido constantemente. En ROB y RNA, este incremento comenzó entre 2000 y 2002. En AA, el aumento notable ha sucedido durante los últimos seis años, especialmente en los dos últimos. El promedio de citas por trabajo es de 10, aunque en RNA asciende a 11. Estos trabajos incluyen tanto la producción nacional como las colaboraciones internacionales.

La *Figura 2* presenta la distribución de las publicaciones según su autoría nacional o internacional. Las colaboraciones internacionales inician en el 2000; constituyen 44% de la producción total y tienden al incremento. En AA, la colaboración supera a la producción nacional durante varios años, alcanzando 55% en el 2022. En ese mismo año, la colaboración en VC es de 64% y en el área genérica de la IA ascendió a 57%. En RNA y ROB la cantidad de colaboraciones se incrementa entre 2% y 5% por año, pero manteniéndose ligeramente por debajo de 50%:

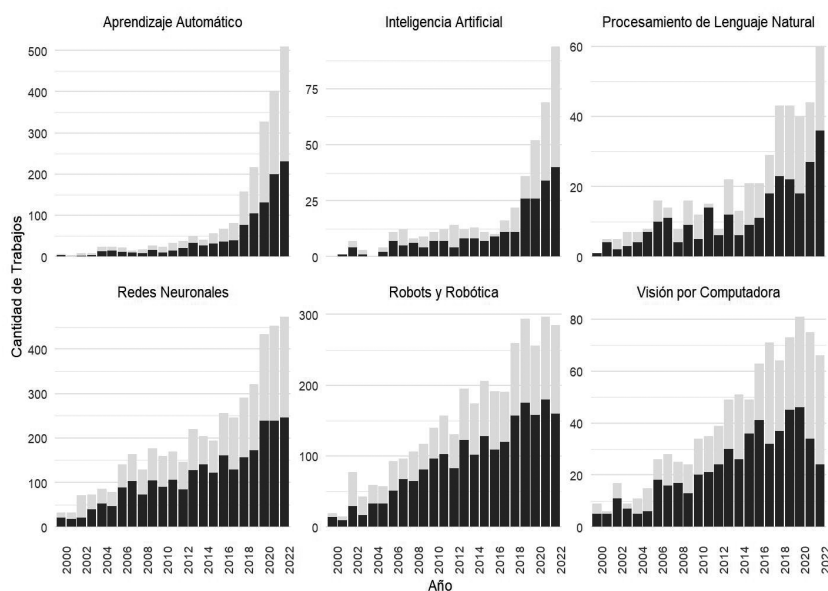


Figura 2. Distribución de la producción mexicana e internacional
 Matrices claros: colaboraciones internacionales. Matrices oscuros: autoría mexicana
 Fuente: elaboración de la autora

La productividad de los autores mexicanos corresponde con la importancia de las colaboraciones internacionales. La *Tabla 3* indica la distribución de autores

por subcampo y tipo de autoría según su institución de filiación. Esta aparece con tasas altas, que van desde 29% en ROB hasta 44% en AA y AP, y una tasa promedio de tres publicaciones por autor. La misma tabla muestra el porcentaje de primeros autores mexicanos, que varía por subcampos. ROB y RNA poseen mayor cantidad de primeros autores, 71% y 63% respectivamente, mientras que áreas más recientes, como AA, PLN y VC tienen más colaboraciones que primeras autorías:

Subcampo	Autores				Trabajos			
	Únicos (total de la base)	Filiación mexicana	Filiación no mexicana	Filiación sin identificar	Total	Primer autor mexicano	Primer autor mexicano (%)	Promedio por autor único
AA y AP	8665	4248	3821	596	2566	1766	42	3
PLN	1474	911	421	142	511	384	42	3
VC	3187	2045	930	212	1012	798	39	3
RNA	10152	6263	3193	142	4955	3955	63	2
ROB	6655	4300	1897	458	3713	3067	71	2
IA	1879	1043	702	134	510	369	35	4
Total general	32012	18810	10964	1684	13267	10339	55	3

Tabla 3. Autores mexicanos e internacionales según el país de filiación institucional
Fuente: elaboración de la autora

No obstante, la *Figura 3* expone un crecimiento significativo de las primeras autorías principalmente en AA, y un aplanamiento en ROB:

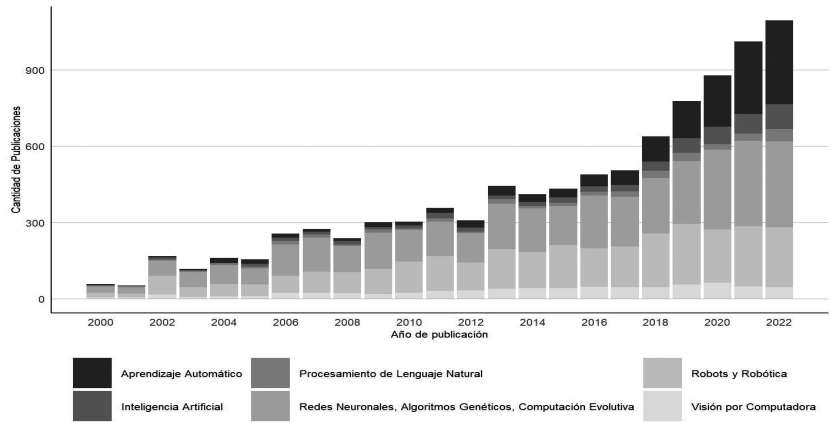


Figura 3. Primeros autores mexicanos por subcampo
Fuente: elaboración de la autora

La cantidad de publicaciones por autor indica una tasa alta de distribución. Tomando en cuenta el promedio de tres trabajos indicado en la *Tabla 4*, se dividió a los autores en ‘prolíficos’ y ‘no prolíficos’. Para determinar el umbral entre ambos se siguió a Garg y Padhi (2000: 360), quienes consideran prolíficos a los autores cuya producción alcanza el doble del promedio y no prolíficos a quienes se encuentran por debajo de esa medida. Por lo tanto, cuando su productividad era de 6 o más artículos se les consideró prolíficos y cuando era menor a 6, se les consideró no prolíficos. Para todos los subcampos, menos el de RNA, los segundos son los que aportan la mayor cantidad de trabajos y citas; sin embargo, debe considerarse la alta cantidad de citas de los autores prolíficos en todos los subcampos, a pesar de ser numéricamente menores en las áreas emergentes. Los subcampos más equilibrados entre autores prolíficos y no prolíficos son los maduros: RNA y ROB.

Subcampo	Tipo de autor	Cantidad de autores	Autores (%)	Cantidad de citas	Citas (%)
AA y AP	Prolíficos	138	3	16766	34
	No prolíficos	4110	97	32744	66
PLN	Prolíficos	21	2	3112	21
	No prolíficos	890	98	11680	79
VC	Prolíficos	33	2	5089	24
	No prolíficos	2012	98	15700	76
RNA	Prolíficos	353	6	74882	51
	No prolíficos	5910	94	71899	49
ROB	Prolíficos	300	7	46894	55
	No prolíficos	4000	93	39132	45
IA	Prolíficos	11	1	698	9
	No prolíficos	1032	99	6965	91

Tabla 4. Distribución de autores por nivel de publicaciones y citas

Fuente: elaboración de la autora

Respecto a la distribución por género, se encontró una diferencia de 24% femenino contra 76% masculino en autorías mexicanas, siendo el área de procesamiento del lenguaje (PLN) donde había una diferencia menor, 29% de autoras. La robótica (ROB) fue el área con una brecha mayor, de 17%. Estos datos son análogos a la distribución de género de los autores internacionales (*Tabla 5*):

Subcampo	Reconocidas	Femeninas	Femeninas (%)
AA y AP	3796	930	25
PLN	797	229	29
VC	1874	476	25
RNA	5220	158	20
ROB	3495	110	17
IA	786	11	25

Tabla 5. Autorías mexicanas por género
Fuente: elaboración de la autora

México está ubicado entre los primeros cuarenta países en producción académica en todos los subcampos de IA, variando desde el puesto 39.º en AA hasta el 23.º en ROB. La *Figura 4* resalta la colaboración con países líderes y con países latinoamericanos. En cuanto a los centros de investigación, destaca la relevancia de universidades, institutos y centros públicos, y de una sola universidad de financiamiento privado. No se han notado especializaciones: todos se dedican a los diferentes subcampos, aunque el Instituto Politécnico Nacional (IPN) y el Cinvestav (Centro de Investigación y de Estudios Avanzados del IPN) reúnen mayor cantidad de investigaciones en ROB y RNA, el Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) en AA, y la UNAM tiene una distribución equitativa en AA, ROB y RNA.

La mayoría de los trabajos no detallan sus fuentes de financiamiento. Los que lo hacen informan a Conahcyt (Consejo Nacional de Humanidades, Ciencias y Tecnologías), IPN y UNAM como los más frecuentes. A nivel internacional, las Fundaciones Nacionales de Ciencia de Estados Unidos y China, así como la Fundación Bill y Melinda Gates son las más destacadas.

Mapeo temático

Esta sección ofrece una visión general de los subcampos de investigación a través de dos enfoques. El primero estudia los conceptos atribuidos por OpenAlex a cada publicación. El segundo muestra un análisis de las palabras presentes en sus títulos.

Para el análisis del desarrollo temático se seleccionaron los cinco primeros conceptos atribuidos a cada artículo que contaran con una frecuencia de repetición de diez o más veces en toda la base y se estudió su progreso. Una visión general de los conceptos con mayor producción en 2022 presenta dos agrupaciones. En la primera están aquellos que evolucionaron a lo largo de las últimas

tres décadas; por ejemplo, las RNA, la teoría del control, la VC y el diseño de robots. En la segunda se incluyen los que progresaron rápidamente en los últimos años: el AA, las redes neuronales convolucionales y el AP. Tanto por su progresión como por la cantidad de la producción reciente, todos los mencionados forman un grupo de conceptos consolidados. La *Figura 5* evidencia los veinte más frecuentes:

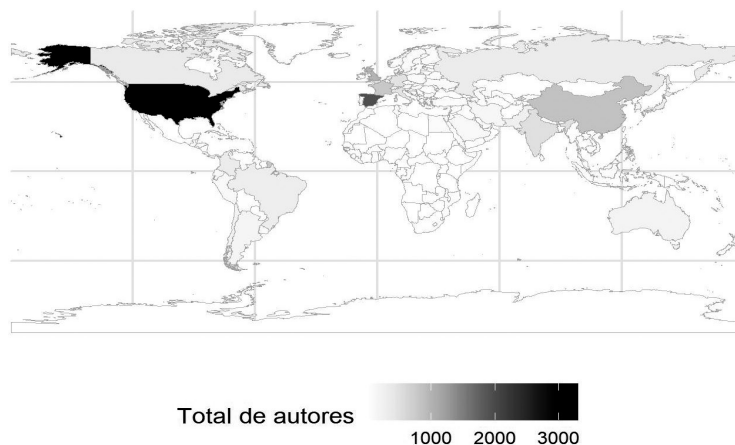


Figura 4. Colaboraciones internacionales de autores mexicanos

Fuente: elaboración de la autora

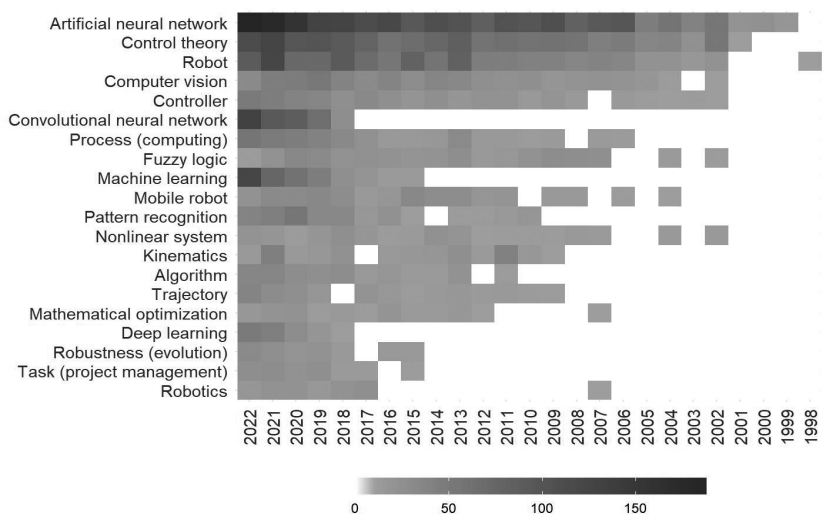


Figura 5. Desarrollo de conceptos consolidados

Fuente: elaboración de la autora

Otro grupo de conceptos evolucionó en los últimos cinco años. Se compone de temas emergentes y perfila un conjunto de áreas de aplicación de la IA; por ejemplo, la medicina, en particular hacia la COVID-19; la ingeniería; la matemática; las finanzas; el procesamiento del lenguaje y del habla y el desplazamiento de robots. Un grupo más reducido lo componen algunos conceptos en declive, como “reconocimiento del habla”, “búsqueda de información”, “ciencia cognitiva”, “heurística”, “educación matemática”, “aprendizaje de refuerzo” y “algoritmos evolutivos”.

El segundo enfoque consiste en un análisis de palabras presentes en los títulos de todos los trabajos. Se fragmentaron los títulos en palabras únicas y se graficó una red de palabras concurrentes con aquellas que aparecían más de diez veces en la base de datos (Callon, Courtial y Laville, 1991: 161). La red resultante consta de 1 661 nodos y palabras, y 123 623 aristas, que representan los trabajos conectados por las mismas palabras. Las estadísticas de la red revelaron un grado medio de 149 conexiones por nodo y una centralidad de grado de 0.9, lo cual indica que cada nodo está conectado con 9% de la red. El diámetro de la red es de 3 y su coeficiente de agrupamiento es de 0.47 que, a su vez, muestra una tendencia a la formación de grupos.

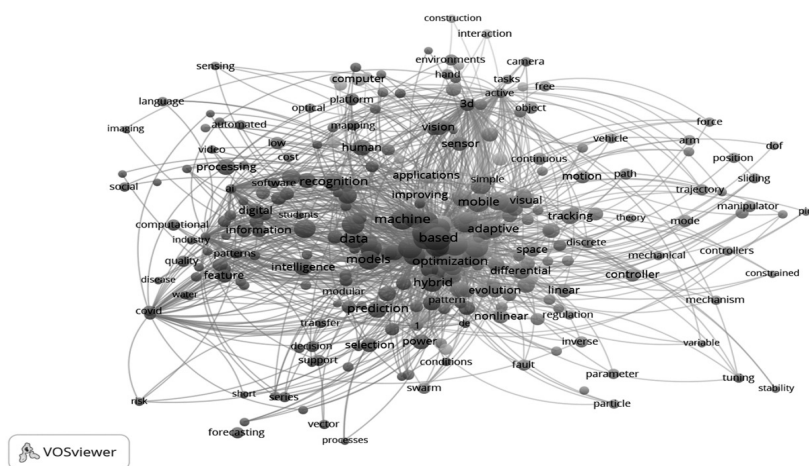
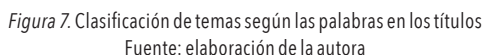


Figura 6. Agrupamiento no supervisado de temáticas
Derecha: robótica. Izquierda: aprendizaje automático. Arriba: visión por computadora
Fuente: elaboración de la autora

La *Figura 6* demuestra el análisis de modularidad, que reveló tres principales agrupamientos: uno vinculado a modelos de aprendizaje automático, reconocimiento de patrones y procesamiento del lenguaje; otro asociado a diseño y estudio de robots; y un tercero relacionado con la visión por computadora. En la periferia, se identificaron términos aplicados al aprendizaje automático como

Utilizando las métricas de centralidad y densidad se clasificaron las palabras en cuatro categorías temáticas: motoras, consolidadas, subdesarrolladas y de nicho, representadas en la *Figura 7* (Callon, Courtial y Laville, 1991: 165-66).



- a) Los temas motores tienen alta centralidad y densidad, siendo fundamentales y bien desarrollados. Es el caso de los aprendizajes automático y profundo, los robots, el reconocimiento de imágenes y los procesos de optimización y clasificación.
- b) Los temas consolidados, con baja centralidad y alta densidad, están bien desarrollados, pero menos integrados en la red. Pueden ser trabajados por grupos o comunidades específicas. Aquí encontramos ciertos temas relacionados con la salud, el reconocimiento facial, la acústica y la robótica aplicada a vehículos.
- c) Los temas subdesarrollados, de baja centralidad y densidad, son periféricos y emergentes. En su especificidad son similares a los consolidados,

pero menos trabajados. Se identifican temáticas de salud, agricultura, movimiento y vehículos.

- d) Los temas de nicho poseen baja centralidad y alta densidad. No se identificaron, pero podrían surgir próximamente si los temas emergentes se consolidan en grupos de investigación específicos.

DISCUSIÓN

Este artículo presentó el primer análisis bibliométrico sobre investigación en IA en México. Su contribución más relevante es haber caracterizado la producción mexicana en inteligencia artificial como un campo de investigación en progreso constante, con subcampos que se comportan diferenciadamente en su desarrollo. ROB y RNA son subcampos consolidados con un progreso construido a lo largo de más de veinte años. AA, PLN y VC son subcampos emergentes, cuya irrupción y crecimiento ocurrió en los últimos cinco años. IA como subcampo incluyó pocos artículos inespecíficos, lo cual indica una especialización importante en torno a la IA. Se destaca que en México existe producción científica para todos los subcampos seleccionados, lo cual manifiesta una base sólida para el desarrollo futuro en el área. Será interesante conocer cómo se aprovechará el conocimiento ya generado.

Los subcampos maduros y emergentes revelaron patrones diferentes de autoría y colaboraciones internacionales. En los primeros, el número de autores ‘prolíficos’ superó al de ‘no prolíficos’ y las colaboraciones internacionales fueron menores a 50%. En los segundos predominaron los autores con menos de seis publicaciones, quienes aportan la mayor cantidad de citas. Las colaboraciones internacionales superaron la mitad de las publicaciones; dichas colaboraciones se dan con países líderes en cada subcampo y con algunos países de América Latina. Los estudios venideros deberán enfocarse en las ventajas y desventajas de la colaboración internacional en términos de infraestructuras y de agendas locales de investigación. La distinción entre autores ‘prolíficos’ y ‘no prolíficos’ posibilita acercamientos que versen en la construcción de trayectorias científicas en cada subcampo.

Las primeras autorías de filiación mexicana aumentaron significativamente en los últimos cinco años para AA y RNA. ROB también ha construido un liderazgo, pero a lo largo de más de dos décadas. Debe prestarse atención a VC que pese a su buen posicionamiento global ha perdido primeras autorías en los últimos años. México está ubicado entre los primeros cuarenta países con mayor producción en todos los subcampos. También debe destacarse que la distribución por género evidencia una brecha importante entre mujeres y varones: el

porcentaje promedio de autoras mexicanas es de 23.5%. Futuras investigaciones deberán ahondar, cualitativamente, en las dificultades para su reducción y en la productividad de las mujeres en el área.

En el mapeo temático encontramos tanto conceptos extensamente desarrollados y trabajados durante las últimas tres décadas —redes neuronales, teoría del control y visión por computadora— como aquellos que irrumpen en los últimos cinco años: el AA, el AP y las redes neuronales convolucionales. Las aplicaciones en diferentes campos, como la medicina, las finanzas, la matemática y el medioambiente, también aparecen en ese período. Sería conveniente que próximamente también se aborde cómo cuadran las agendas de investigación locales con las recomendaciones de políticas públicas orientadas al bien social (Gómez-Mont *et al.*, 2020: 13).

Los subcampos motores que emergen del mapeo temático son la robótica, el aprendizaje automático y la visión por computadora que, a su vez, mantiene vínculos estrechos con los dos primeros. La partición en tres agrupamientos temáticos interroga la división de seis subcampos adoptada de Girasa (2020: 33-39) como criterio único. De esta forma, futuros acercamientos podrán indagar en criterios de clasificación situados y adecuados para investigar áreas de conocimiento. Las particularidades señaladas abonan al uso de las mediciones bibliométricas para la construcción de indicadores de evaluación para las tecnologías emergentes (Amaro-Rosales y Robles-Belmont, 2020: 17), aunque los estudios cualitativos venideros podrán brindar más información sobre las dinámicas develadas en este artículo.

CONCLUSIONES

La literatura revisada coincide en que, a pesar del constante crecimiento de la producción científica en IA, faltan estudios bibliométricos tanto globales (Loan, Bashir y Nasreen, 2021: 29), como mexicanos (López-Martínez y Sierra, 2021: 184). Este artículo aporta al conocimiento del área pues diferencia subcampos con diferentes comportamientos en la producción mexicana.

Los trabajos previos en torno a la producción bibliométrica mexicana anticipan la relevancia de la producción científica en IA destacando su producción a nivel global, identificando a tres autores del área como los más citados y resaltando su importancia en la producción de uno de los principales institutos de investigación en ciencias de la computación (Arencibia-Jorge *et al.*, 2023: 10; Lancho-Barrantes y Cantú-Ortiz, 2019: 506; Ruiz León, 2018: 17). El presente trabajo va más lejos al precisar el lugar global ocupado por la investigación mexicana. Segunda en América Latina, los diferentes subcampos sostienen desde el

puesto 39.º en aprendizaje automático, hasta el 23.º en robótica. Este lugar destacado precisa la afirmación general de la importancia de la producción mexicana en robótica realizada por Cornejo *et al.* (2023: 1435).

Este estudio logró identificar progresos en las temáticas de investigación de los últimos años. En el estudio global de Loan, Bashir y Nasreen (2021: 31) los conceptos asociados a AA y AP no aparecen entre los más frecuentes, mientras que en el de López-Martínez y Sierra (2021: 189), desarrollado sobre producción mexicana hasta 2017, aparecen marginalmente. Puede distinguirse un cambio de paradigma donde el aprendizaje automático y el profundo realizan una rápida curva ascendente y predominan en la investigación de los últimos cinco años. Estos hallazgos reafirman la tesis de Amaro-Rosales y Robles-Belmont (2020: 17) acerca de la necesidad de modelos innovadores, así como de evaluaciones dinámicas y constantes de las tecnologías disruptivas y emergentes.

Agradecimientos

Artículo financiado por Conahcyt, Estancias Postdoctorales por México. Este trabajo se realizó igualmente con el respaldo del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIT) de la UNAM IN302623 “Indicadores sobre la ciencia y la tecnología en el contexto de la Ciencia Abierta”.

REFERENCIAS

- Amaro-Rosales, Marcela, y Eduardo Robles-Belmont. 2020. “Medir la innovación en el contexto de las tecnologías emergentes y convergentes: algunas reflexiones metodológicas”. *PAAKAT: Revista de Tecnología y Sociedad* 10 (18): 1-20.
<https://doi.org/10.32870/Pk.a10n18.415>
- Arencibia-Jorge, Ricardo, Ibis Anette Lozano-Díaz, José Luis Jiménez-Andrade y Humberto Carrillo-Calvet. 2023. “A Multidimensional Approach to Mexican Scientific Output from 2010-2019”. *Transinformação* 35: 1-14.
<https://doi.org/10.1590/2318-0889202335e237320>
- Aria, Massimo, Corrado Cuccurullo, Trang Le y June Choe. 2023. “openalexR: Getting Bibliographic Records from ‘OpenAlex’ Database Using ‘DSL’ API”. V. 1.3.1.
<https://cran.r-project.org/web/packages/openalexR/index.html>
- Bastian, Mathieu, Sebastien Heymann y Mathieu Jacomy. 2009. “Gephi: An Open Source Software for Exploring and Manipulating Networks”. *Proceedings of the Third International AAAI Conference on Weblogs and Social Media* 3 (1): 361-62.
<https://doi.org/10.1609/icwsm.v3i1.13937>
- Bredahl, Laura. 2022. *The Current and Evolving Landscape of Bibliometric Tools and Technologies*. Chicago: American Library Association.
- Callon, Michel, Jean Pierre Courtial y Françoise Laville. 1991. “Co-Word Analysis as a Tool for Describing the Network of Interactions between Basic and Technological Research: The Case of Polymer Chemistry”. *Scientometrics* 22 (1): 155-205.
<https://doi.org/10.1007/BF02019280>

- Cornejo, José, Salvador Barrera, Cesar Herrera-Ruiz, Francisco Gutiérrez, Marcelo Casasnovas, Leonardo Kot, Miguel Solís, *et al.* 2023. "Industrial, Collaborative and Mobile Robotics in Latin America: Review of Mechatronic Technologies for Advanced Automation". *Emerging Science Journal* 7 (4): 1430-58.
<https://doi.org/10.28991/ESJ-2023-07-04-025>
- Demografix ApS. s.f. Genderize.io. Consultado el 26 de noviembre de 2023.
<https://genderize.io/>
- Donthu, Naveen, Satish Kumar, Debmalya Mukherjee, Nitesh Pandey y Weng Marc Lim. 2021. "How to Conduct a Bibliometric Analysis: An Overview and Guidelines". *Journal of Business Research* 133: 285-96.
<https://doi.org/10.1016/j.jbusres.2021.04.070>
- Eck, Nees Jan van, y Ludo Waltman. 2019. *VosViewer*. V. 1.6.20. Universidad de Leiden.
<https://www.vosviewer.com/>
- Garg, Kailash, y Pitambara Padhi. 2000. "Scientometrics of Prolific and Non-Prolific Authors in Laser Science and Technology". *Scientometrics* 49 (3): 359-71.
<https://doi.org/10.1023/A:1010529405152>
- Girasa, Rosario. 2020. *Artificial Intelligence as a Disruptive Technology: Economic Transformation and Government Regulation*. Cham: Palgrave McMillan.
- Gómez-Mont, Constanza, Claudia May Del Pozo, Cristina Martínez Pinto y Ana Victoria Martín del Campo Alcocer. 2020. *La inteligencia artificial al servicio del bien social en América Latina y el Caribe. Panorámica regional e instantáneas de doce países*. Washington D.C.: Banco Interamericano de Desarrollo.
<https://doi.org/10.18235/0002393>
- Kaplan, Andreas, y Michael Haenlein. 2019. "Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence". *Business Horizons* 62 (1): 15-25.
<https://doi.org/10.1016/j.bushor.2018.08.004>
- Lancho-Barrantes, Bárbara, y Francisco Cantú-Ortiz. 2019. "Science in Mexico: A Bibliometric Analysis". *Scientometrics* 118 (2): 499-517.
<https://doi.org/10.1007/s11192-018-2985-2>
- Loan, Fayaz Ahmad, Bisma Bashir y Nahida Nasreen. 2021. "Applied Artificial Intelligence: A Bibliometric Study of an International Journal". *COLLNET Journal of Scientometrics and Information Management* 15 (1): 27-45.
<https://doi.org/10.1080/09737766.2021.1938742>
- López-Martínez, Roberto, y Gerardo Sierra. 2021. "State of Research on Natural Language Processing in Mexico — A Bibliometric Study". *Journal of Data, Information and Management* 3 (3): 183-95.
<https://doi.org/10.1007/s42488-021-00051-5>
- OurResearch. 2022. OpenAlex. Consultado el 26 de septiembre de 2023.
<https://openalex.org/>
- Priem, Jason, Heather Piwowar y Richard Orr. 2022. "OpenAlex: A Fully-Open Index of Scholarly Works, Authors, Venues, Institutions, and Concepts". Prepublicación ArXiv.
<https://doi.org/10.48550/arXiv.2205.01833>
- Queiroz, Gabriela de, Colin Fay, Emil Hvitfeldt, Os Keyes, Kanishka Misra, Tim Mastny, Jeff Erickson, David Robinson y Julia Silge. 2023. "tidytext: Text Mining Using 'dplyr', 'ggplot2', and Other Tidy Tools". V. 0.4.2.
<https://cran.r-project.org/web/packages/tidytext/index.html>

- RStudio Team. 2020. *RStudio: Integrated Development for R*. V. 2023.06. Desarrollado por Posit. <http://www.rstudio.com/>
- Ruiz León, Alejandro. 2018. "Análisis de la producción y colaboración científica de instituciones: el caso del Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la Universidad Nacional Autónoma de México (UNAM)". *Redes. Revista Hispana para el Análisis de Redes Sociales* 29 (1): 1-19. <https://doi.org/10.5565/rev/redes.755>
- Sued, Gabriela Elisa. 2024a. "Dataset for Bibliometric Analysis of Mexican Scientific Production on Artificial Intelligence". *Zenodo*, 10 de febrero de 2024. <https://doi.org/10.5281/zenodo.10642462>
- Sued, Gabriela Elisa. 2024b. "Bibliometria_de_la_IA_enMexico.R". https://github.com/ObservatorioCtsMx/OpenAlex-con-R/blob/main/Bibliometria_de_la_IA_enMexico.R
- Vergara Villegas, Osslan Osiris, Manuel Nandayapa, Juan Humberto Sossa Azuela y Félix Agustín Castro Espinoza. 2021. "Editorial: A Brief Panorama of Artificial Intelligence in Mexico". *International Journal of Combinatorial Optimization Problems and Informatics* 12 (3): 1-7. <https://www.ijcopi.org/ojs/article/view/272>
- Wickham, Hadley, Davis Vaughan, Maximilian Girlich, Kevin Ushey, *Posit* y *PBC*. 2023a. "tidyr: Tidy Messy Data". V. 1.3.1. <https://cran.r-project.org/web/packages/tidyr/index.html>
- Wickham, Hadley, Romain François, Lionel Henry, Kiril Muller y Davis Vaughan. 2023b. "dplyr: A Grammar of Data Manipulation". V. 1.1.4. <https://dplyr.tidyverse.org>
- Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, Dewey Dunnington, Teun van den Brand y *RStudio*. 2023c. "ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics". V. 3.5.1. <https://CRAN.R-project.org/package=ggplot2>

Para citar este texto:

- Sued, Gabriela Elisa. 2024. "La producción científica mexicana en inteligencia artificial: un análisis bibliométrico". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 38 (100): 87-105. <http://dx.doi.org/10.22201/iibi.24488321xe.2024.100.58893>