Journal of Physics:
Conference Series

PURPOSE-LED PUBLISHING™

PAPER • **OPEN ACCESS**

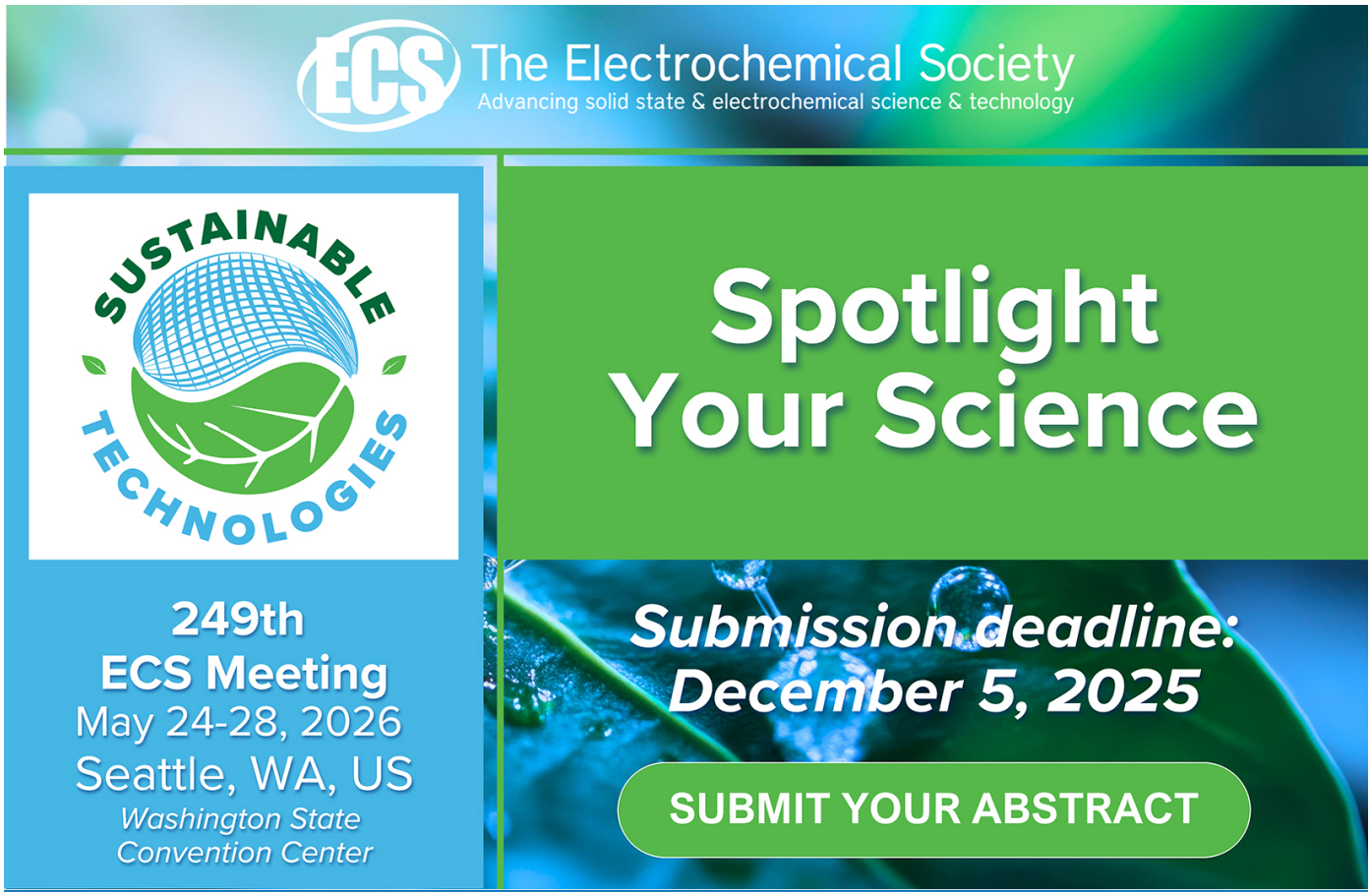
Artificial intelligence theory: a bibliometric analysis

To cite this article: E Romero-Riaño *et al* 2021 *J. Phys.: Conf. Ser.* **2046** 012078

View the article online for updates and enhancements.

You may also like

- Bibliometrics and altmetrics on halitosis: exploring the top 100 most-cited papers
Ana Paula Carvalho, Karolina Skarlet Silva Viana, Fernando Oliveira Costa et al.
- Bibliometric analysis of global research trends on biomimetics, biomimicry, bionics, and bio-inspired concepts in civil engineering using the Scopus database
Naim Sedira, Jorge Pinto, Isabel Bentes et al.
- A bibliometric analysis of wildlife governance in relation to African elephant conservation and community participation
Tanyaradzwa Mundoga, Walter Musakwa and Nelson Chanza



This content was downloaded from IP address 8.243.64.131 on 17/10/2025 at 20:24

Artificial intelligence theory: a bibliometric analysis

E Romero-Riaño[1], D Rico-Bautista[2], M Martinez-Toro[3], Y Medina-Cárdenas[2], N Rico-Bautista[2]
[1] Universidad Industrial de Santander, Bucaramanga, Colombia
[2] Universidad Francisco de Paula Santander, Seccional Ocaña, Colombia
[3] Universidad Autónoma de Bucaramanga, Bucaramanga, Colombia

E-mail: dwricob@ufpso.edu.co, efren.romero@saber.uis.edu.co

Abstract. This article analyzes the evolution of artificial intelligence research from the bibliometric perspective. Bibliometrics, as one sub-field of scientometric, is defined as the use of statistical methods for analyzing publication data. A total of 8334 papers were identified divided into two periods, 2010-2014 and 2015-2019. To recognize the trends of artificial intelligence research, bibliometric analysis and social network analysis are combined to judge the current situation and development trends. The results revealed the keywords with the highest occurrence and those with the strongest linkage strength from cluster analysis. Moreover, the study of artificial intelligence was found to be an active field of growth, and the words that control the knowledge area were identified. The results of this study will facilitate the understanding of the progress and trends in artificial intelligence for researchers interested in understanding its evolution.

1. Introduction

In the results generated at the international level [1], a literature review from 1990 to 2017, 97% of the literature focuses on the internet of things (IoT) and artificial intelligence (AI); more specifically, 55% of the literature focuses on AI and 42% of the studies focus on IoT. This observation allows identifying that the main technologies used are essentially IoT and AI [2-6].

To analyze the trends of artificial intelligence research [7-11], bibliometric analysis and social network analysis are combined to analyze and judge the current situation and development trends [12,13]. Social network analysis is a powerful model of social structure integration [14,15], , which has some important formal aspects of analyzing networks and their participants [16]. Important tools in the face of the large growth of academic papers related to artificial intelligence [17,18].

When combined, bibliometrics and socialization network analysis can explore the development of artificial intelligence; influential journals; the most important researchers; keywords, among other characteristics [19]. Using bibliometric and social network methods, this study produced a quantitative and qualitative analysis of the most popular topics in artificial intelligence in recent years. The present article focuses on the analysis of written texts, published in journals indexed within WoS Clarivate, considered as the most influential and reliable database [16].

The document is structured, the first section introduces; the second section presents the methodology, the third section, bibliometric analysis and a social network analysis around artificial intelligence are presented, and finally, the results and suggestions.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

2. Methodology and materials

The selection of the papers arises as a result of the search expression: subject: ("artificial intelligence") or subject: ("AI"). Period: 2010-2019. Indexes: sci expanded, Ssci, A&Hci, Esci within the WoS Core Collection. A total of 8334 documents were identified, which, for the study, will be divided into two periods, 2010-2014 and 2015-2019. For the period corresponding to the years 2015-2019, a total of 6151 are identified; for the period 2010-2014, the number of documents is 2183 records.

The bibliometric analysis supports the quantitative analysis of scientific knowledge by using mathematical, statistical, and visual approaches to generalize research patterns and trends in scientific publications [16]. For the development of this article, the methodology adapted from Table 1 is presented. This combines methods and tools of network visualization and social network analysis. These methods are complementary and contribute to the qualitative and quantitative approaches.

Table 1. Summary of implementation of the article's methodology.

Phase	Summary
Research question	What is the semantic structure of AI research?
Identification of sources	Web of Science (WoS)
Metadata extraction	Periods 2015-2019 (6151) and 2010-2014 (2183). Batches of 500 complete records with cited references.
Display of test networks	Keyword cooccurrence network visualization (2010-2019). Keyword network overlay map (2010-2019).
Data mining and record cleaning	Thesaurus construction for debugging 34 duplicate tags
Definition of units of analysis	Keywords indexed by the database (Keyword Plus)
Definition of types of bibliometric analyses	Co-word analysis (Cooccurrence of indexed keywords).
Network map visualization (normalization, distribution, and clustering).	Standard VOSviewer displays parameters for normalization, distribution, and clustering. Minimum cluster size: 20 nodes.
Estimation of network centrality indicators.	Indicators are calculated as follows: Complete network: (density) Nodes (centrality, proximity, and intermediation)
Comparative AI analysis of the summaries	Overlay map of terms extracted from document abstracts.

3. Results and discussions

The WoS search yielded the metadata used to perform the bibliometric analysis of co-occurrence of keywords using the VOS viewer software and derive a map in which the structure of the field of knowledge is visualized. Cooccurrence is frequently used in scientometric analysis due to its multiplicities of purposes, such as co-authorship and co-citation [20,21]. In the case of co-occurrence of keywords, the words that have been used together in the sample are analyzed and from there the characteristics of the structure of the field of knowledge are inferred. The sample of articles to be analyzed presents 8192 keywords, to which, as a simplification and stratification criterion for the construction of the map, the minimum of 20 occurrences was defined, obtaining 134 keywords with 3434 connections that meet this principle, which is represented on the maps as circles with their labels, which according to their occurrence, determine their size and which are connected by lines that evidence the strength of the link between terms according to their thickness [20,21], see Figure 1.

From the results obtained from the number of occurrences of the keywords, Table 2 shows the 22 words with the highest number of occurrences in the sample, headed by "models" and "systems" with 692 occurrences each. The concept "models" shows its importance, following the definition of the concept proposed by Pidd [22], as a representation of a reality conceived for a particular purpose, the conceptual basis of artificial intelligence.

In the same way, the word "system" implies the vision of an interconnected whole [23,24], which is intended to be represented using models and the application of artificial intelligence techniques. The words "algorithms" (510 occurrences), "neural networks" (462 occurrences), and "genetic algorithms"

(194 occurrences), have participation as instructions that guide the computational steps for a specific purpose. The terms "classification" (342 occurrences), "optimization" (312 occurrences), "prediction" (274 occurrences), "recognition" (132 occurrences), "identification" (116 occurrences), "selection" (116 occurrences) and "diagnosis" (113 occurrences), are present as the main applications of artificial intelligence techniques.

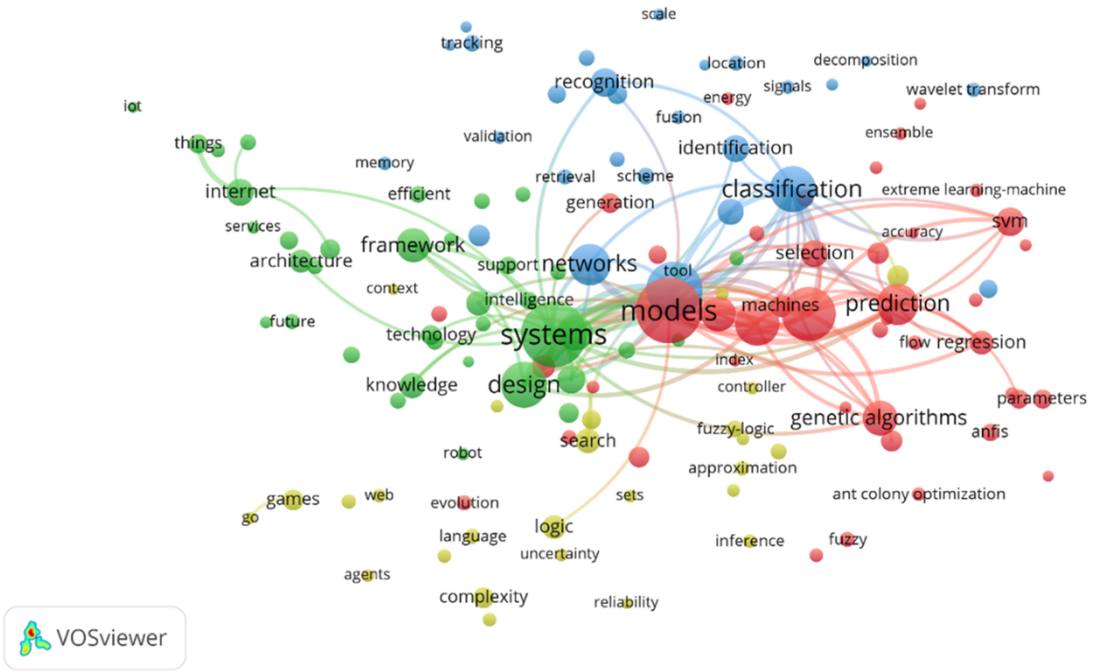


Figure 1. Keyword cooccurrence network map.

Table 2. Cooccurrence of keywords.

Number	Keyword	Occurrence	Number	Keyword	Occurrence
1	Models	692	12	Genetic Algorithms	194
2	Systems	692	13	Framework	188
3	Algorithms	510	14	Recognition	132
4	Neural-Networks	462	15	SVM	131
5	Classification	342	16	Management	127
6	Design	334	17	Internet	117
7	Optimization	312	18	Identification	116
8	Prediction	274	19	Selection	116
9	Networks	272	20	Diagnosis	113
10	Performance	200	21	Knowledge	103
11	Ai	199	22	Information	100

Figure 2 provides information corresponding to the average year in which the term has been used as a keyword. This Figure 2 gives an orientation concerning the validity of the terms and their evolution over time using a classification on a color scale that goes from white to purple, with the terms represented in white having an older average publication value and in purple the most recent ones. The time window analyzed is from 2000 to 2019, in which it is observed that some of the terms with the highest occurrence ("models" and "systems") are represented in white, which has two connotations, the first, that these terms have not been used recently in scientific publications, and the second that the terms have been used since long ago to identify scientific publications, which affects their average.

The terms such as "optimization", "prediction", "design", among others, have been used in the academic spectrum for several decades, showing the maturity of these topics. From the map, it is possible to extract terms such as "challenges", "health care", "deep", "images", "scale", "signals", "validation", which show an average publication value of 2017, and the term "IoT" which presents an average publication value of 2018, which shows them as the most recent topics or emerging topics that are catching the interest of researchers.

The crossword analysis matrix in Table 3 shows the existence and strength of links between the 10 keywords with the highest co-occurrence strength; Table 3 shows the total linkage strength of each of the keywords, which indicates the total number of connections that the word has with the other words within its knowledge network, for example: "models" has a total linkage strength of 1543, "systems" has a total linkage strength of 1510, "neural networks" has a total linkage strength of 1082. By relating the terms, we can see the strength between them, which indicates the number of times they appear together in documents of the analyzed sample, which offers a particular view of the knowledge field.

The term "models" shows the highest cooccurrence link strength value for the terms "systems" (114), "neural networks" (85), "prediction" (67), "networks" (37), and "performance" (43), highlighting the trend towards the representation of systems and their usefulness in the development of network analysis, seeking to predict results and improve network performance. The term "systems" shows the highest cooccurrence linkage strength value for the terms "algorithms" (74), "design" (53), and 'optimization" (57), confirming its character of the representation of structures, using algorithms, for design and optimization purposes.

The socio-centric analysis is performed with the data values for a degree, density, and distance that Ucinet has normalized, the joint word matrix was imported, and treatment was applied to the data by generating a new matrix using the "Transform>Dichotomize" function. This treatment generates, as a result, a focus on the existence or not between AI topics. The generated matrix is used as input for subsequent analyses. In addition, the function "Network>Complete Network and Cohesion>General Density" was applied. Based on this, the density of the co-word network is estimated. The value obtained corresponds to 0.385 with a standard deviation of 0.487. This value shows that taking as a model the network or word system of Figure 1 and Figure 2, the AI research domain is still a developing area of research and presents opportunities for further study.

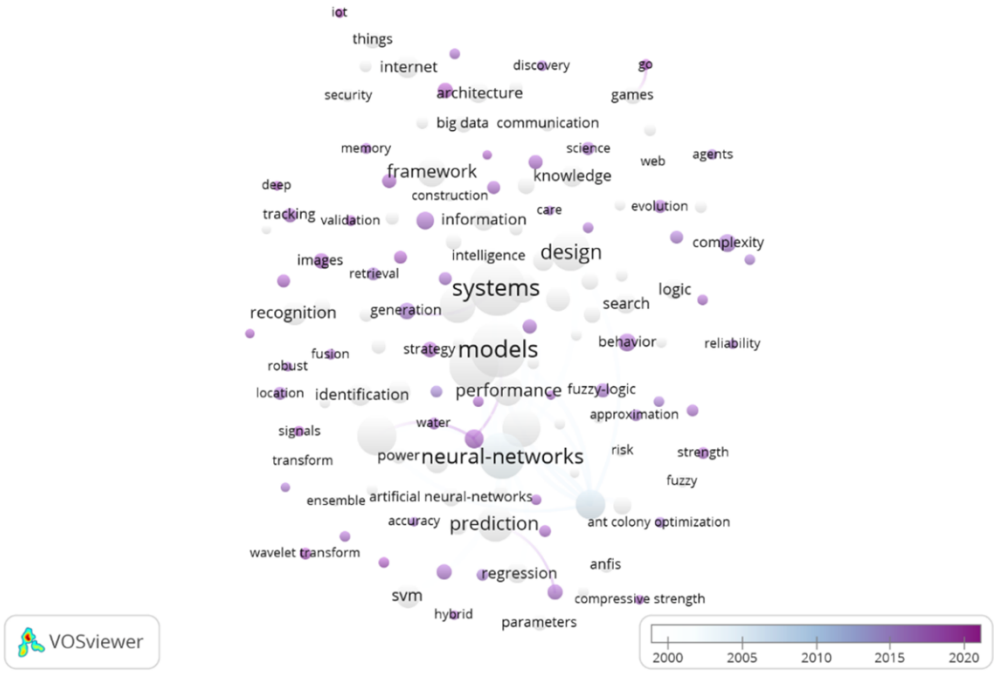


Figure 2. AI research terms network overlay map.

Table 3. Matrix of crossword analysis.

	M	S	N-N	A	C	D	Pr	O	N	Pe
M	1543	114			45					
S	114	1510					45			
N-N			1082					28	0	
A				994						0
C	45				811	0				0
D					0	727	0			
Pr		45				0	714		0	
O								681		0
N			0				0		514	0
Pe				0	0			0	0	473

M = Models; S = Systems; N-N = Neural-Networks; A = Algorithms; C = Classification
D = Design; Pr = Prediction; O = Optimization; N = Networks; Pe = Performance

The topic of "micro-worlds" networks offers a structural perspective in parallel to other existing perspectives, such as quantitative studies of social network. Specifically, a micro-world is defined as a network in which the typical distance between two nodes grows proportionally to the logarithm of the number of nodes "n" in the network. These micro-worlds can be classified according to two independent structural characteristics, namely the clustering coefficient and the average distance between nodes or geodesic distance. Based on this background, a distance-based micro-world analysis is performed using the function "Network>Cohesion>Geodesic distance". It is shown that the estimated score for the average distance within the co-word networks in IA is 1.615 with a standard deviation of 0.487. Consistent with the finding of Swanson [23,24], when the average distance is less than 6, there is a "micro-world" characteristic. This finding can be interpreted as an indication that the study of AI is a growing field of the asset.

The analysis of centrality was carried out using the indicators of the degree of centrality, closeness, and intermediation. The degree of centrality suggests that nodes that have more connections in a network may have a greater influence on the network. For the estimation of the degree of centrality, the function "Network>Centrality>Degree" was implemented. It is identified that the problems most frequently addressed by AI are optimization (0.797) and prediction (0.805), for performance estimation (0.767) using neural networks (0.902), see Table 4. As a complement, the function "Network>Centrality>Closeness" was implemented. Closeness is calculated as the inverse of the geodesic distance between nodes and shows that certain AI research topics are referred to or used simultaneously within the published papers. Based on the indicators in Table 4, it is identified that the words with the closest proximity coincide with the words with the highest degree. Based on this, it is inferred that this group of words shows the research core of this discipline.

Table 4. Summary of centrality measures of the most influential topics in IA.

Word	Standardized grade	Proximity	Interrelation
Models	1.0	1.0	3.937
Systems	0.985	0.985	3.811
Algorithms	0.940	0.943	3.355
Neural-Networks	0.902	0.911	2.901
Networks	0.857	0.875	2.525
Optimization	0.797	0.831	2.079
Prediction	0.805	0.836	2.002
Performance	0.767	0.811	1.759

4. Conclusions

The results showed that there is an increasing amount of research on the topic in recent years. From the map, terms such as "challenges", "health care", "deep", "images", "scale", "signals", "validation", which show an average publication value of 2017, and the term "IoT" which presents an average publication value of 2018, which shows them as the most recent topics or emerging topics that are catching the interest of researchers. The concept "models" shows its importance as a representation of a reality conceived for a particular purpose, the conceptual basis of artificial intelligence. The term "models" turns out to be the one with the highest value of the strength of co-occurrence links. When the average distance is less than 6, there is a "micro-world" characteristic.

This finding can be interpreted as an indication that the study of AI is a growing field of the asset. In the degree and closeness analyses, the nodes design (2.438), classification (2.285), framework (2.036) emerges with the highest values of "interrelation". This means that the identified terms played an important role in the artificial intelligence study and that they have the highest control over the research topics in the whole network. It is necessary to continue with the bibliometric analysis of other technologies such as the internet of things, big data, and cloud computing. Achieving at the end a complete and integrated analysis of the trends around the smart university.

References

[1] Rjab A B, Mellouli S 2018 Smart cities in the era of artificial intelligence and internet of things: literature review from 1990 to 2017 *International Conference Proceeding Series. Association for Computing Machinery* (New York: ACM) p 1
[2] Garfield E 1972 Bibliometric analysis as a tool in journal evaluation *Science* **1972(178)** 471
[3] Pritchard A 1969 Statistical bibliography or bibliometrics *Journal of Documentation* **25(4)** 348
[4] White H D, Mccain K W 1998 Visualizing a discipline: an author co-citation analysis of information science *Journal of the American Society for Information Science* **49(4)** 327
[5] Small H 1973 Co-citation in the scientific literature: a new measure of the relationship between two documents *Journal of the American Society for Information Science* **24(5)** 265
[6] Narin F 1987 Bibliometric techniques in the evaluation of research programs *Sci. Public. Policy* **14(2)** 99
[7] Jusoh M F, *et al.* 2021 An overview of the internet of things (IoT) and irrigation approach through bibliometric analysis *IOP Conf. Ser: Earth. Environ. Sci.* **756(1)** 012041:1
[8] Rico Bautista D, Medina Cardenas Y, Coronel Rojas L, *et al*. 2021 Smart university: key factors for an artificial intelligence adoption model *Advan. and Applicat. in Comput. Science, Electronics and Industrial Engineering* (Quito: Springer) p 153
[9] Liu Z, Yin Y, Liu W, *et al.* 2015 Visualizing the intellectual structure and evolution of innovation systems research: a bibliometric analysis *Scientometrics* **103(1)** 135
[10] Effendi D *et al*. 2021 Bibliometric analysis of scientific literacy using VOS viewer: analysis of science education *Journal of Physics: Conference Series* **1796(1)** 012096:1
[11] Rico Bautista D, Medina Cardenas Y, Areniz Arevalo Y, *et al.* 2020 Smart university: big data adoption model *2020 9th Intern. Confer. on Software Process. Improvement* CIMPS (Sinaloa: IEEE) p 52
[12] Ziyuan L *et al.* 2020 The Development trend of artificial intelligence in cyberspace security: a brief survey *Journal of Physics: Conference Series* **1486(2)** 022047:1
[13] Martinez Toro G, Ariza Zabala G C, Rico Bautista D, *et al.* 2019 Human computer interaction in transport, a systematic literature review *Journal of Physics: Conference Series* **1409(1)** 012002:1
[14] Gil Quintana J, Santoveña Casal S, Riaño E R 2021 Realfooders influencers on Instagram: From followers to consumers *Int. J. Environ. Res. Public. Health* **18(4)** 1624
[15] Galeano Barrera C, Bellón Monsalve D, Zabala Vargas S, *et al.* 2018 Identificación de los pilares que direccionan a una institución universitaria hacia un smart-campus *Rev. Inve. Desa. E. Innova.* **9** 127
[16] Huang M, Wang Z, Chen T 2019 Analysis on the theory and practice of industrial symbiosis based on bibliometrics and social network analysis *J. Clean. Prod.* **213** 956
[17] Lei Y, Liu Z 2019 The development of artificial intelligence: a bibliometric analysis, 2007-2016 *Journal of Physics: Conference Series* **1168(2)** 022027:1
[18] Wang Z, Liao H 2020 Towards the eco-design of artificial intelligence and big data applications: a bibliometric analysis of related research *Conf. Ser. Mater. Sci. Eng.* **806(1)** 012039:1

[19] Romero Riaño E, Guarín-Manrique L, Dueñas Gómez M, *et al.* 2019 Reference framework for capabilities development in agricultural innovation system *DYNA* **86(210)** 23
[20] Tu W, *et al.* 2020 Application analysis of machine learning in fault diagnosis: A bibliometric analysis *Journal of Physics: Conference Series* **1629(1)** 012020:1
[21] Wang L 2021 Key technology and application of computer artificial intelligence recognition based on machine Vision *Journal of Physics: Conference Series* **1982(1)** 012112:1
[22] Man Z, *et al.* 2020 Exploring research fronts and topics of big data and artificial intelligence application for cultural heritage and museum research *Conference Series: Materials Science and Engineering* **806** 012036
[23] Niu Y, *et al.* 2021 The development tendency of artificial intelligence in command and control: a brief survey *Journal of Physics: Conference Series* **1883(1)** 012152:1
[24] González Sarmiento E, Roa Perez J, Ortiz Ospino L 2021 Big data and artificial intelligence in the development of industry 4.0; a bibliometric analysis *Conf. Ser. Mater. Sci. Eng.* **1154(1)** 012008:1