Scientometric Analysis of the Research Paper Output on Artificial Intelligence: A Study

H. Mumtaj Begum
Librarian, Anna Centenary Library, Chennai, Tamil Nadu, India
E-mail: mumtajh@gmail.com

Abstract- **The purpose of this study was to measure the number of contributions and highlight the contributions made by the researchers in the field of Artificial Intelligence and published on the Web of Science database ten years during 2011-2021 using scientometric analysis. Data were interpreted by using software such as Bibexcel, Vosviewer, and tabulated using MS Excel. The results indicated that 26822 papers were published during 2011 - 2021 and the highest number of publications 10712 (39.94%) was produced in 2020. The trends in multi-authored papers have tremendously increased (87.32%) compared to (12.68%) single-authored papers. The relative growth rate (1.46) and degree of collaboration (0.87) is noted significantly and the highest no of papers (18.97%) was contributed by the collaboration of three authors and source wise most of the records were published an article 20741. It also noted that the value of the highest degree of collaboration was (0.89) in 2020. The Zhang Y author (111) contributed more numbers of papers in the domain of Artificial Intelligence with USA (6491) being the country producing more research papers followed by China, the UK, and the India. More than 1600 papers had been published in Model, and Prediction. The study inferred that the rate of growth is relation by the year wise publications of Artificial Intelligence research.**
Keywords: **Artificial Intelligence, Expert System, Machine Learning, Scientometrics, Authorship Pattern**

I. INTRODUCTION

A. Scientometrics

Scientometrics is the field of study which concerns itself with measuring and analyzing scholarly literature. Scientometrics is a sub-field of bibliometrics. Major research issues include the measurement of the impact of research papers and academic journals, the understanding of scientific citations, and the use of such measurements in policy and management contexts. In practice there is a significant overlap between scientometrics and other scientific fields such as information systems, information science, science of science policy, sociology of science, and metascience. Critics have argued that over-reliance on scientometrics has created a system of perverse incentives, producing a publish or perish environment that leads to low quality research.

B. Artificial Intelligence

Artificial Intelligence is the ability of a computer program to learn and think. John McCarthy coined the term 'Artificial Intelligence' in the 1950s. He said, 'Every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions, and concepts, solve kinds of problems now reserved for humans, and improve themselves.'

C. Benefits of Artificial Intelligence

It enables multi-tasking and eases the workload for existing resources. AI enables the execution of hitherto complex tasks without significant cost outlays. AI operates 24x7 without interruption or breaks and has no downtime. AI augments the capabilities of differently able individuals.

II. REVIEW OF LITERATURE

Mustak, M., *et al.,* (2021) have explained this study, through the application of natural language processing, machine learning, and statistical algorithms, we examine extant literature in terms of its dominant topics, diversity, evolution over time, and dynamics to map the existing knowledge base. Tensalient research themes emerge: (1) understanding consumer sentiments, (2) industrial opportunities of AI, (3) analyzing customer satisfaction, (4) electronic word-of-mouth–based insights, (5) improving market performance, (6) using AI for brand management, (7) measuring and enhancing customer loyalty and trust, (8) AI and novel services, (9) using AI to improve customer relationships, and (10) AI and strategic marketing. The scientometric analyses reveal key concepts, keyword co-occurrences, authorship networks, top research themes, landmark publications, and the evolution of the research field over time. With the insights as a foundation, this article closes with a proposed agenda for further research.

Darko, A. *et al.,* (2020) have described this paper, the first comprehensive scientometric study appraising the state-of-the-art of research on AI-in-the-AECI is presented. The science mapping method was used to systematically and quantitatively analyze 41,827 related bibliographic records retrieved from Scopus. The results indicated that genetic algorithms, neural networks, fuzzy logic, fuzzy sets, and machine learning have been the most widely used AI

methods in AEC. Optimization, simulation, uncertainty, project management, and bridges have been the most commonly addressed topics/issues using AI methods/ concepts. The primary value and uniqueness of this study lies in it being the first in providing an up-to-date inclusive, big picture of the literature on AI-in-the-AECI. This study adds value to the AEC literature through visualizing and understanding trends and patterns, identifying main research interests, journals, institutions, and countries, and how these are linked within now-available studies on AI-in-the- AECI. The findings bring to light the deficiencies in the current research and provide paths for future research, where they indicated that future research opportunities lie in applying robotic automation and convolutional neural networks to AEC problems. For the world of practice, the study offers a readily- available point of reference for practitioners, policy makers, and research and development (R&D) bodies. This study therefore raises the level of awareness of AI and facilitates building the intellectual wealth of the AI area in the AEC industry.

Zhong, B., *et al.,* (2019) examines this study reviews ontology research mainly published in the Scopus database from 2007 to 2017 with the combination of scientometric analysis and critical review. Scientometric analysis objectively visualized the research status quo while a critical review was used to identify the research themes and challenges of ontology research in the construction industry. The results identified a large network of co-authors in this field to understand collaboration relationships. Over half the papers (53%) were published by the following three countries: the United States, the United Kingdom, and Canada. The top co-occurring keywords were "project management" at which ontology facilitates knowledge management and information retrieval. When the time factor was taken into consideration, keywords naturally evolved from "project management", and "knowledge management" to "building information modeling", and "compliance control" with the successful adoption of information techniques in the construction industry. Four research themes were identified with the combination of cluster analysis and critical review: "Domain ontology", "Industry foundation classes", "Automated compliance checking", and "Building information modeling".

Gupta, B. M., & Dhawan, S. M. (2018) examines the world output in artificial intelligence research, a total of 1,52,655 publications, as seen from Scopus database, covering the period during 2007-16. The top 10 countries of the world in artificial intelligence research accounted for 74.32 per cent global publication share. Individually their global share varied from 3.68 per cent to 19.46 per cent, with China accounting for 19.46 per cent global share, followed by the USA (17.96 %), India (6.37 %), and the U.K. (6.33 %), etc. The paper also examines publications output by India in artificial intelligence research. India cumulated a total of 9730 publications in 10 years during 2007-16, registered an annual average growth rate of 27.45 per cent, averaged

citation impact to 2.76 citations per paper, and contributed 10.34 per cent share of its total country output as international collaborative publications during 2007-16. Computer science accounted for the largest publication share (86.99 %), followed by engineering (30.69 %), mathematics (15.95 %), biochemistry, genetics & molecular biology (4.66 %), and several other disciplines. The top 10 organizations and 10 authors together accounted for 19.31 per cent and 2.71 per cent national publications share respectively and 29.78 per cent share and 6.85 per cent national citation share respectively during 2007-16. Top 10 journals accounted for 15.45 per cent share of the country output appearing in journal medium (1650 papers). India accounted for 24 highly cited papers, averaging to 162.46 citations per paper. These 24 highly cited papers involved the participation of 109 authors from 70 organizations, published in 15 journals.

Shrivastava, R., & Mahajan, P. (2016) explore the study analyzes the research output of India in the field of artificial intelligence using scientometric analysis techniques. The data were collected manually using the Scopus database at the end of July 2015. Publications in the field of artificial intelligence research in India from 1968 to 2014 were retrieved. It was found that a total of 6,529 papers were published in the field of artificial intelligence in India during that time period. The research output has grown considerably since 2004, and the last 2 years have witnessed a large publications output from the field. The average citation per paper of this data set is 3.06. The average number of authors per paper is three. "Artificial Intelligence" was found to be the most popular keyword, followed by "Algorithms." A total of 12.64% of the papers have been published with international collaboration. Anna University was found to be the leader in research productivity. It was found that the IITs played a major role in the field of artificial intelligence research in India.

III. OBJECTIVES OF THE STUDY

1. To find the Year-wise distribution of articles. (2011-2021)
2. To study the pattern of growth of publications during the period.
3. To find out the document type of the publications.
4. To analyze the authorship pattern, degree of collaboration.
5. Relative growth rate and doubling time.
6. To know the country-wise distribution of the publications.

IV. METHODOLOGY

Topic: Artificial Intelligence Period: 2011-2020. The total number of articles retrieved was 26822. The retrieved articles were analyzed using Bibexcel software. Further analysis was done using MS Excel and VOS Viewer software.

TABLE I DATA AND INTERPRETATIONS YEAR-WISE DISTRIBUTION OF ARTIFICIAL INTELLIGENCE

Sl. No	Year	Records	Percent	Sl. No	Year	Records	Percent
1	2011	669	2.49	7	2017	1685	6.28
2	2012	749	2.79	8	2018	3057	11.40
3	2013	828	3.09	9	2019	6039	22.51
4	2014	939	3.50	10	2020	10712	39.94
5	2015	989	3.69	Total		26822	100
6	2016	1155	4.31				

Table I shows the distribution of articles published on Artificial Intelligence research during 2011-2020(10 years). It can be observed that from 2011 to 2020, there was a frequent change of increase in the publications. The maximum number of articles was published in 2020 with 10712 (39.94%) records, the next highest publication could be observed in the year 2019 with 6039 (22.51%) records. The least number of articles was observed in the year 2011 with 669 (2.49%) records.

TABLE II ANNUAL GROWTH AND CITATIONS

Sl. No	Year	Records	Cumulative number of Records	%	Article Citations	H-index	Average Citation per item
1	2011	669		2.49	17747	64	27.55
2	2012	749	1418	2.79	18677	65	26.59
3	2013	828	2246	3.09	19130	64	24.4
4	2014	939	3185	3.50	20152	64	23.29
5	2015	989	4174	3.69	23201	69	25.36
6	2016	1155	5329	4.31	30576	74	28.82
7	2017	1685	7014	6.28	41359	90	27.84
8	2018	3057	10071	11.40	52287	105	21.81
9	2019	6039	16110	22.51	57470	97	13.38
10	2020	10712	26822	39.94	49012	106	11.80
Total		26822		100			

Table II reveals that the total number of papers published and citations received during 2011 to 2020 in the "Artificial Intelligence". The highest numbers of 10712 (39.94%) records were published in 2020 and a minimum number of 669(2.49%) records were published in 2011. The growth of publications was not in constant during the study period.

When considering the citation, there are totally 329611 citations were received during the study period. Out of 329611, there are 57470 citations were received on 2019 which is highest and 106 citations were received in 2020 during the study period.

TABLE III RELATIVE GROWTH RATE AND DOUBLING TIME

Year	Total Publications	Cumulative No. of Publications	W1*	W2*	R(a) (W2W1)	MeanR(a) 1-2*	DoublingTime Dt(a)	Mean Dt (a)1-2
2011	669	-	6.50	-	-	0.86	-	
2012	749	1418	6.61	7.25	0.64		1.08	
2013	828	2246	6.71	7.71	1.00		0.69	0.56
2014	939	3185	6.84	8.06	1.22		0.56	
2015	989	4174	6.89	8.33	1.44		0.48	
2016	1155	5329	7.05	8.58	1.53	1.21	0.45	
2017	1685	7014	7.42	8.85	1.43		0.48	
2018	3057	10071	8.02	9.21	1.19		0.58	0.59
2019	6039	16110	8.70	9.68	0.98		0.70	
2020	10712	26822	9.27	10.19	0.92		0.75	
Total	26822							

Table III depicts that Relative Growth Rate and Doubling Time of total publications. Though the publication from 2011 (669 records) to 2020 (10712 records), the relative growth rate was 0.64 in 2012 which increased up to 1.53 in 2016. A mean relative growth rate of 1.46 could be deduced for the study period. The mean doubling time during the period 2011-2020 was 0.57 and for 2011 -2020 it was decreased from 1.08 to 0.45.

*R (1-2) = Mean Relative Growth Rate over the Specified Period interval;
*W1= log w1 (Natural log of the initial number of publications)
*W2 = log w2 (Natural log of the initial number of publications)
*T2-T1= the unit Difference between the initial time and final time.
*R (a) = Relative Growth Rate per unit publication per unit of time (Year)

Formula

R (1 – 2) = W2 – W1 / T2 – T1Dt(a) = 0.693 / R(a)

Where,

R (1-2) = Mean relative growth rate for the specified period.
W1 = log w1 (Natural log of the initial number of publications)
W2 = log w2 (Natural log of the initial number of publication
T2-T1 = the unit difference between the initial and the final time.
R (a) = relative growth rate per unit publication per unit of time.

TABLE IV ANNUAL RATIO OF GROWTH AND ANNUAL GROWTH RATE

Year	Total Publications	ARoG*	AGR*
2011	669		
2012	749	1.11	0.11
2013	828	1.10	0.10
2014	939	1.13	0.13
2015	989	1.05	0.05
2016	1155	1.16	0.16
2017	1685	1.45	0.45
2018	3057	1.81	0.81
2019	6039	1.97	0.97
2020	10712	1.77	0.77
Total	26822		

ARoG* - Annual Ratio of Growth AGR* - Annual Growth Rate

The exponential growth rate of publications in Artificial Intelligence research during 2011-2020 could be seen in table IV The maximum growth rate (0.97), and ARoG 1.97 was found in the year 2019 (6039 records), followed by 0.81 and ARoG 1.81 in the year 2018 (3057 records). The least exponential growth rate (0.05), and ARoG 1.05 was observed in the year 2015 (989 records). It shows the variations in growth rate from 2011 to 2020.

TABLE V SINGLE AUTHOR VERSUS MULTI-AUTHORS

Sl. No	Authors	Publications	Percentage
1	Single Authors	3401	12.68
2	Multi Authors	23421	87.32
	Total	26822	100

Table IV shows the contributions of a single author and multiple authors of Artificial Intelligence research during the period 2011-2020. Single authors had contributed 3401 (12.68%) publications and multiple authors had contributed 23421 (87.32%) publications out of 26822 publications. Hence, it was inferred that the majority of the articles were published by multiple authors.

TABLE VI AUTHORSHIP PATTERN

Sl. No	Authors	Publications	Percentage
1	One Authors	3401	12.68
2	Two Authors	4460	16.63
3	Three Authors	5089	18.97
4	Four Authors	4426	16.50
5	Five Authors	3131	11.67
6	Six Authors	1992	7.44
7	Seven Authors	1145	4.27
8	Eight Authors	792	2.95
9	Nine Authors	566	2.11
10	Ten Authors	433	1.61
11	Above Ten Authors	1387	5.17
	Total	26822	100

Table VI shows the authorship pattern in Artificial Intelligence research during 2011-2020. The most number of papers (5089) had been brought out by the collaborative work of Three authors followed by two authors (4460), four authors (4426), one authors (3401), five authors (3131), above ten authors (1387), seven authors (1145), eight authors (792), nine authors (566), ten authors (433).

Table VII shows the authorship pattern of publications by year based on collaborative research.

TABLE VII YEAR-WISE AUTHORSHIP PATTERN

Year	1	2	3	4	5	6	7	8	9	10	10+	Total
2011	116	163	174	119	52	19	8	6	2	3	7	669
2012	110	170	202	149	58	23	14	8	4	3	8	749
2013	100	184	213	174	80	33	18	12	5	3	6	828
2014	124	194	225	203	95	43	23	11	8	3	10	939
2015	148	206	257	168	113	44	22	6	9	1	15	989
2016	157	241	301	185	112	66	34	21	10	10	18	1155
2017	312	342	337	266	178	87	62	25	27	13	36	1685
2018	454	529	585	480	372	219	101	78	63	40	136	3057
2019	789	937	994	965	731	485	265	191	148	131	403	6039
2020	1091	1494	1801	1717	1340	973	598	434	290	226	748	10712
	3401	4460	5089	4426	3131	1992	1145	792	566	433	1387	26822

TABLE VIII DEGREE OF COLLABORATION

Year	Single Author (NS)	Multiple Authors (NM)	Total (NM + NS)	Degree of Collaboration C=NM/NM+NS
2011	116	553	669	0.82
2012	110	639	749	0.85
2013	100	728	828	0.87
2014	124	815	939	0.86
2015	148	841	989	0.85
2016	157	998	1155	0.86
2017	312	1373	1685	0.81
2018	454	2603	3057	0.85
2019	789	5250	6039	0.86
2020	1091	9621	10712	0.89
Total	3401	23421	26822	0.87

Table VIII amounts to calculate the degree of collaboration in Artificial Intelligence research by using the formula given by (K. Subramanyam, 1982) which was,

C = NM / NM + NS

Where,
C = the degree of collaboration

NM = no. of multi-authored papers
NS = no. of single-authored papers

The degree of collaboration varied from 0.81 to 0.89 for the duration of the study. The average mean value was 0.87 suggesting the increase of collaborative research in Artificial Intelligence.

TABLE IX RANKING OF AUTHORS BASED ON NO. OF PUBLICATIONS

Rank	Authors	Records
1	Zhang Y	111
2	Liu Y	106
3	Zhang J	95
4	Li J	94
5	Wang Y	92
6	[Anonymous]	91
7	Wang J	90
8	Li Y	84
9	Wang L	78
10	Lee S	72

The top 10 rank prolific authors were recognized in Artificial Intelligence research. They had published 70 or more papers during 2011-2020. The identified top 10 ranked authors had brought about 913 (3.40%) papers. An Zhang Y person became the most productive author who contributed 111 articles, followed by Liu Y with 106 articles.

TABLE X PUBLICATION TYPE BASED ON H-INDEX

h-index	Unit	Citation sum within h-core	All citations	All articles
158	Article	57576	304077	20741
125	Review	34272	81464	2906
44	Editorial Material	3645	9785	1605
11	Letter	215	452	180
8	News Item	327	392	92
6	Meeting Abstract	50	281	1129
3	Book Review	15	21	89
2	Correction	9	22	66
1	Retraction	4	4	3
1	Biographical-Item	3	3	8
1	Reprint	1	2	2
0	Poetry	0	0	1

Table X reveals the type of document based on H-index used for publication during the period 2011-2020. It could be seen that the most of the publications and H-index were in Articles with 20741 records, H-index 158, followed by Review with 2906 records and 125 H-index, Editorial Material with 1605 records and 44 H-index, and Letter, News Item, Meeting Abstract, Book Review, Correction, Retraction, Biographical-Item, Reprint, and Poetry are followed.

TABLE XI KEYWORD-WISE DISTRIBUTION

Rank	Keywords	Records
1	Model	1663
2	Prediction	1661
3	Classification	1449
4	System	1209
5	Neural-Networks	1090
6	Optimization	1040
7	Algorithm	1015
8	Artificial-Intelligence	973
9	Design	952
10	Performance	949

Table XII had clearly shown that the most frequent keyword. We have selected only most (Top Ten) used keywords for analysis. The result that the most productive keyword "Model" has been used in 1663 records, followed by the word "Prediction" used in 1661 records, "Classification" used in 1449 records, "System" with 1209 records, "Neural-Networks" with 1090 records, "Optimization" with 1040 records, 'Algorithm' with 1015 records and others are bellowed 1000 records.

TABLE XII MOST PRODUCTIVE COUNTRIES

Rank	Country	Records
1	USA	6491
2	China	5674
3	UK	2545
4	India	1502
5	Spain	1383
6	Iran	1360
7	Germany	1342
8	Canada	1253
9	Australia	1154
10	South Korea	1100

Table XII gives the distribution of Artificial Intelligence publications by the country during 2011- 2020. USA had ranked top among the countries in terms of publications with 6491 records, followed by China 5674, UK 2545 records, India 1502 records. and Spain, Iran, Germany, Canada, Australia, and South Korea had bellowed 1500 records. India had contributed 1502 and ranks 4th in the top 10 countries.

TABLE XIII LANGUAGE-WISE DISTRIBUTION

Rank	Language	Records
1	English	26282
2	German	190
3	Spanish	106
4	French	63
5	Portuguese	45
6	Chinese	29
7	Russian	23
8	Polish	14
9	Turkish	13
10	Hungarian	12
11	Croatian	8
12	Czech	7
13	Japanese	7
14	Italian	6
15	Korean	4
16	Slovak	3
17	Dutch	3
18	Icelandic	2
19	Ukrainian	1
20	Unspecified	1
21	Lithuanian	1
22	Slovene	1
23	Swedish	1

V. FINDINGS AND CONCLUSION

After the analysis of data derived from the Web of Science database for the period 2011- 2021(10 Years), here are presented following interesting facts, findings or we can say in more formal word results of the study that is given below.

1. The distribution of articles by year displayed that a total of 26822 publications were published in Artificial Intelligence during 2011-2021 (10 years). The maximum number of articles 10712 (39.94%) were published in the year 2020.
2. The Maximum number of 57470 citations received in the year 2019, and least of 17747 citations were received in 2011. The numbers of citations have been gradually decreased year by year.
3. There could be observed an increasing and decreasing trend in the relative growth rate and doubling time data. The mean relative growth rate for the period 2011-2020 was 1.46. The mean doubling time was 0.57.
4. Multiple authors i.e., 23421 (87.32%) had contributed more than the single authors i.e., 3401(12.68%).
5. The authorship pattern indicated that the maximum number of papers was published by the collaborative work of three authors 5089 and the least number of papers published by group of ten authors with 433 records.
6. The Degree of Collaboration varied from 0.81 to 0.89 during the period of the study. High collaborative research could be seen from the mean value of 0.87.
7. From the prolific author's list, it could be seen that Zhang Y author was the most productive author with 111 records
8. The distribution of publications by document type was found mostly in articles with 20741 publications and 158 H-index during the study period.
9. The most common keyword used by the researcher was "Model" with 1663 of the total records.
10. The top 10 distribution of articles by the country during 2011-2020 ranked USA at the top among all the countries with 6491 publications.
11. The distribution of ORL literature by language showed that the common scholarly communication was in the English language with 26282 articles.

REFERENCES

[1] Mustak, M., Salminen, J., Plé, L., & Wirtz, J. (2021). Artificial intelligence in marketing: Topic modeling, scientometric analysis, and research agenda. *Journal of Business Research*, *124*, 389-404.
[2] Darko, A., Chan, A. P., Adabre, M. A., Edwards, D. J., Hosseini, M. R., & Ameyaw, E. E. (2020). Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities. *Automation in Construction*, *112*, 103081.
[3] Zhong, B., Wu, H., Li, H., Sepasgozar, S., Luo, H., & He, L. (2019). A scientometric analysis and critical review of construction related ontology research. *Automation in Construction*, *101*, 17-31.
[4] Gupta, B. M., & Dhawan, S. M. (2018). Artificial Intelligence Research in India: A Scientometric Assessment of Publications Output during 2007-16. *DESIDOC Journal of Library & Information Technology*, *38*(6)
[5] Shrivastava, R., & Mahajan, P. (2016). Artificial intelligence research in India: a scientometric analysis. *Science & Technology Libraries*, *35*(2), 136-151.
[6] What is Artificia - AI Definition & Application (intellipaat.com) Scientometrics – Wikipedia.