Artificial Intelligence (AI) based Cyber Security Solution during the Period of 2013-2022 : a bibliometric study

Debalina Mukherjee

Librarian, RCC Institute of Information Technology, Kolkata

Dr. Subal Kumar Barui

Deputy University Librarian, University of Calcutta

Abstract

The integration of Artificial Intelligence (AI) and cyber security holds immense promise for shaping the future of digital defence. The inclusion of cyber security in AI, machine learning research is an emerging trend for academic research. This paper presents a bibliometric analysis of AI based cyber security solutions published in the Scopus data base during the past decade (2013-2022). Data was retrieved on 15th July 2023 by using an advanced search query. It has been revealed from the study that the research is in the early stage and after the COVID -19 the work exhibits its acceleration. India is ranked first among the nations, followed by the USA; in terms of publishing papers on AI based cyber security. Anticipating further advancement in the coming years, this study provides valuable insights into the current state of AI driven cyber security research. Future research could investigate optimal approaches to seamlessly integrate AI and cyber security concepts, ensuring that the next generation of professionals will be well equipped to overcome the problems related to cyber threats.

Keywords: Artificial intelligence, AI, Bibliometric analysis, Bibliometric techniques, Cyber security, Cyber security solutions, Scopus

1. Introduction

In today's internet era, there are diverse research topics in cyberspace. AI has the potential to change the field of cyber security. It can offer the flexibility and precision necessary to adeptly counter cyber threats' continuously changing and evolving landscape. Several conferences, workshops, and journals focus on this research area.

The study seeks to provide insights into the evolution of research, identify key contributors, and map out the intellectual structure of this interdisciplinary domain by examining the patterns of scholarly publications, citations, collaborations, and emerging trends.

Due to the rapid development of information technology and dynamic changes of threats in cyberspace, AI and machine learning play significant roles in cyber security.

2. Literature review

Bibliometricis a measurement process that is used to evaluate and predict the trends of development of science and technology using mathematical, statistical analysis. AI study is highly inter disciplinary because a wide range of journals have been published on AI research. Among them, most of the research ersuse bibliometrics to explore the use and spread of cyber security and AI in their scientific works.

Bircan and Salah (2022) described the Big Data techniques and their computational approaches in social sciences by using bibliometrics. The articles were indexed between 2015 and 2020 in Social Sciences Citation Index (SSCI). Talan (2021) published a paper on artificial intelligence in education indexed in the Web of Science database by using bibliometric analysis. VOS viewer software was used to analyse and visualise all this information. Shukla and Gochhait (2020) studied approximately 2184 records by using Web of Science database and gave a complete idea of the development of cyber security as a research field. Sharma et. al. (2023) presented an extensive bibliometric analysis of cyber security and cyber forensic research published in Web of Science during 2011-2021. Cheng, and Wang (2012) revealed several issues by doing a bibliometric study on AI related publications.

3. Significance of the study

The existing literature review highlighted studies that covered diverse aspects of cyber security and AI research. The present study focuses on a specific intersection - the application of AI in cyber security in a specific period. It addresses a research gap and provides valuable insights into the collaborative efforts between these two domains.

4. Objectives

The objectives of the study are:

- To prepare the chronological distribution of literature and its progress rate
- To identify the contribution of highly active authors in publications and their publications overtime.
- To show the journals with the highest number of publications
- To trace the most active countries and organizations
- To locate the main research are as
- To identify the co-occurrence of author key words

5. Methodology

Data was retrieved on 15th July 2023 using an advanced search query. Microsoft Excel and Biblioshiny (R Tools) software were used for executing science mapping analysis. VOS viewers were used for data visualisation.

Figure 1 shows the inclusion and exclusion criteria for selection of papers (n).

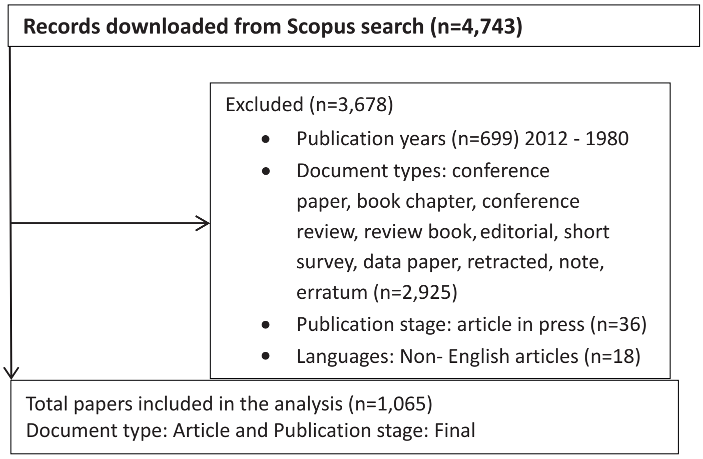
Records downloaded from Scopus search (n=4,743)
Excluded (n=3,678)
Publication years (n=699) 2012 - 1980
Document types: conference paper, book chapter, conference review, review book, editorial, short survey, data paper, retracted, note, erratum (n=2,925)
Publication stage: article in press (n=36)
Languages: Non- English articles (n=18)
Total papers included in the analysis (n=1,065)
Document type: Article and Publication stage: Final

Figure 1: Selection of papers in the Scopus data base

Table 1 shows the main information about extracted data for the research.

Table 1: Main information about data

Description	**Results**
Time span	Year 2013To2022
Sources (Journals, Books, etc)	391
Total Documents	**1065**
Annual Growth Rate %	97.05
Document Average Age	2.38
Average citations per doc	23.12
References	51244
Document contents	
Keywords Plus (ID)	4698
Author's Keywords (DE)	2609
Authorship Pattern	
Total Authors	3378
Authorsofsingle-authoreddocs	65
Authors Collaborations	
Single - authored docs	70
Co - Authors per Doc	3.84
International co-authorships %	30.05
Document Types	
Total articles	1065

6. Limitations

- The study is based on literature on AI based cyber security indexed in the Scopus data base during 2013-2022.
- Scopus permits to export of 2,000 records at a time but does not allow to splitting of the selected collection into multiple downloads
- The language is limited to English

7. Analysis

There are a total of 1065 articles published in 391 sources contributed by 3378 authors during the year 2013-2022.

7.1 *Distribution, citations and progress rate*

Table 2 shows year wise distribution of publications and citations which displays the trends in publications. In the past 10 years total of 1065 papers were published containing an average of 56.14 citations per article and the average total citation per year was 8.19. Most articles were published in 2022 which is more than 400 times higher than that of the year 2013. The most citations per article were made in the year 2015. The most average total citations per year were made in the year 2015. The most citable year was 2013.

Table 2: Year wise distribution of literature and citations

Year	Average no. of citations per article	To tal articles	% of articles published	Average total citations per year	Citable years
2013	3	1	0.09	0.27	11
2014	22.4	5	0.47	2.24	10
2015	207.5	10	0.93	23.06	9
2016	153.46	13	1.22	19.18	8
2017	25.17	18	1.69	3.6	7
2018	54.52	60	5.63	9.09	6
2019	39.58	139	13.05	7.92	5
2020	36.26	156	14.66	9.06	4
2021	13.7	215	20.20	4.57	3
2022	5.82	448	42.06	2.91	2
Total	56.14	1065	100	8.19	

The progress rate of the literature is shown in figure 2. The data shows an increasing trend in research with the upwards looping of publications. The research output has shown a consistent annual increase with a significant surge observed in 2019 and reaching its peak in 2022 with the highest recorded growth.

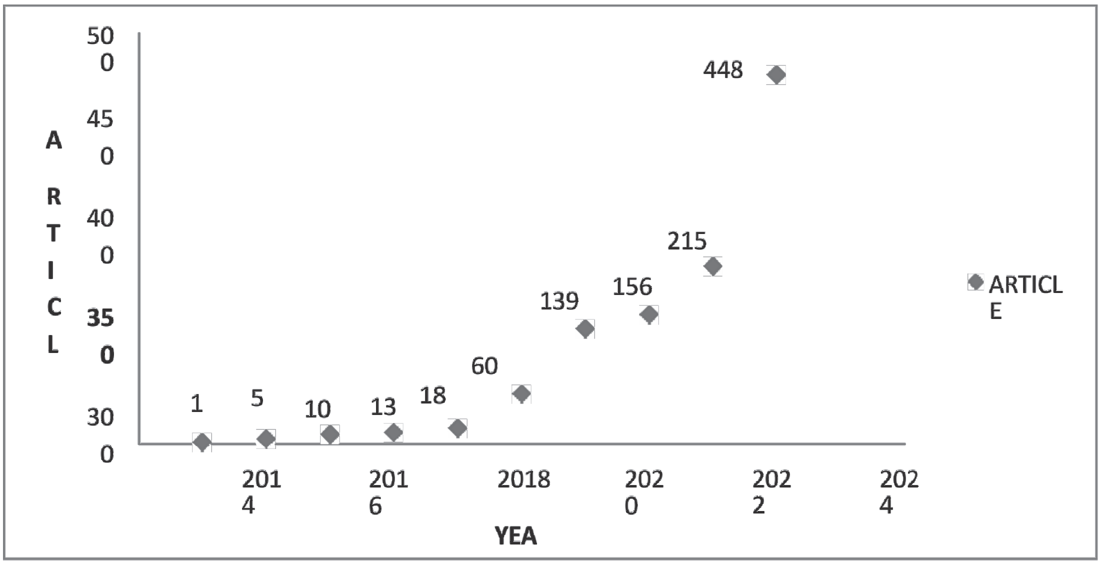

Figure 2: Progress rate of the literature

Rodgers developed the well-known "S Curve" of innovation diffusion theory; innovation diffusion typically displays a distinctive S-shaped curve. Figure 2 shows that diffusion starts very slowly. By accelerating, the curve has reached a certain stage which is called the "critical mass". Innovation speed will be picked up until the research reaches its end point.

7.2 Most active authors

Table 3: Most active authors (top ten)

Author	h_index	g_index	m_index	Total citations	No. articles	Pub. year started
Zhang, J	9	11	1.8	651	11	2019
Liu, Y	7	11	1.4	201	11	2019
Xiang, Y	7	9	1.4	481	9	2019
Alazab, M	6	7	1.2	1433	7	2019
Ferrag, M	6	7	1.5	626	7	2020
Kozik, R	6	8	0.6	219	8	2014
Naeem, H	6	6	1.2	421	6	2019
Pan, L	6	6	1.2	244	6	2019
Pawlicki, M	6	8	1.2	182	8	2019
Chookk, R	5	6	1	274	6	2019

Table 3 lists the top ten authors in order of publications. They have an almost equal number of publications. Among the top ten authors, LIU Y and ZHANG J. have published more than ten papers. Only one author has nine papers among all the top ten authors. Most of the authors have started to publish from the year 2019. Only one author named KOZIKR, published from the year 2014.

The most active author LIU Y's notable achievements include i) the thesis on the machine learning approach to detect cyber-attacks, ii) a novel approach to the detection of cyber-attacks taking inventory of the practical application of information granules, iii) the performance evaluation of intrusion detection algorithm and detection of attacks. Researchers can read following research as each author has a certain direction in their works.

7.3 *Journals with highest publications*

Bradford's law can be used to determine the primary journals on a particular topic. This law demonstrates how scholarly writing is dispersed in journals.

Table 4: Top ten journals (publications number, frequency, percentage and citation)

Source	Rank	Articles	Cumulative frequency	% of Articles	Citations	Zone
IEEE Access	1	97	97	9.107981	1472	Zone 1
Sensors	2	46	143	4.319249	366	Zone 1
Computers and security	3	32	175	3.004695	297	Zone 1
International journal of advanced computer science and applications	4	31	206	2.910798	256	Zone 1
Electronics (Switzerland)	5	24	230	2.253521	218	Zone 1
Computers materials and continua	6	22	252	2.065728	192	Zone 1
Future generation computer	7	16	268	1.502347	189	Zone 1
Applied sciences (Switzerland)	8	15	283	1.408451	184	Zone 1
Computers and electrical engineering	9	14	297	1.314554	179	Zone 1
Expert systems with applications	10	12	309	1.126761	169	Zone 1
TOTAL 309 29.01408						

1065 research papers were distributed in 391 source journals. Table 4 provides the list of key journals. IEEE Access is found to have a 9.1 percent of total literature with the highest number of publications (97 papers) and citations (1472 records). It is to be noted that the top 10 journals listed in table 4 collectively account for 29% of the total number of articles published across all mentioned journals. Notably, IEEE Access (n=97, 9.11%) and Sensors (n=46, 4.31%) emerged as the leading journals in terms of paper publications, followed by the Journal of Computer and Security (n=32, 3%), International Journal of Advanced Computer Science and Applications (n=31, 2.91%).

Bradford's law states that the primary journals of a field are those that publish 33 percent of the published articles in that field. As a result, the mentioned ten journals have been designated as the key journals.

7.4 *Most active countries and organisations*

Table 5: Top ten (10) country's production

Country	Articles	% of Articles
India	191	17.93
United States	187	17.55
China	137	12.86
Saudi Arabia	94	8.82
United Kingdom	90	8.45
Australia	75	7.04
South Korea	51	4.78
Canada	40	3.75
Italy	39	3.66
Turkey	39	3.66
Total	943	88.5

Table 5 displays the distribution of 943 papers focused on AI in cyber security highlighting the leading nations based on publication. India leads in research activity with 191 articles closely followed by the USA with 187 articles. China is in 3rd position regarding the publication of research articles on AI in Cyber security. A total of 943 articles (88.5% articles) out of 1065 are produced by the top ten active countries. Indian scholars are conducting the most thorough research on AI for cyber security.

A total of 160 institutions contributed to 1065 publications. Table 6 shows here the 10 institutions with the highest number of publications. It can be observed that USA and India have more research works on cyber security. Among the top ten institutions, Saudi Arabia is well deserved leader.

Table 6: Top ten notable organisations

Affiliation	Country	Articles
King Abdulaziz university	Saudi Arabia	16
Prince Sattam Bin Abdulaziz University	Saudi Arabia	15
King Saud University	Saudi Arabia	13
Swinburne University of Technolo gy	Australia	12
Deakin University	Australia	12
Qatar University	Qatar	11
King Khalid University	Saudi Arabia	11
Taif University	Saudi Arabia	10
University of New South Wales	Australia	10
Amrita School of Engineering	Karnataka	9

7.5 *Main research areas*

The top ten research areas in the field of artificial intelligence based cyber security are shown in figure 3. Researchers in computer science (40%) and engineering (27%) fields are primarily interested in research. Rests of the fields are below 10%. Other disciplines like Mathematics, Social sciences, Physics are also involved in research on AI for cyber security.

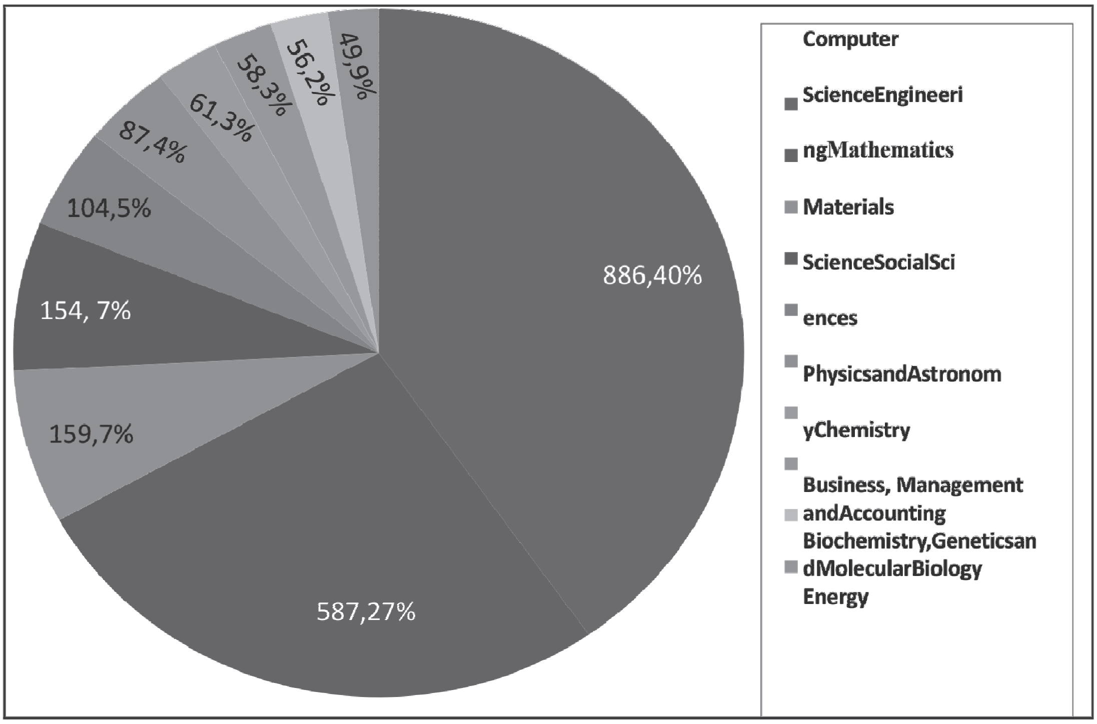

Figure 3: Main research areas

7.6 *Author - keyword co-occurrence analysis*

Visual analysis of author-keyword co-occurrence can provide insight into topics and research trends. 2617 distinct keywords in all are uncovered in the articles. It is found that 21 keywords are most often utilised which indicates that these keywords are directly associated with research. Anomaly detection, artificial intelligence, big data, classification, cyber security, cyber-security, cybersecurity, data mining, deep learning, feature selection and internet of things are the top ten terms. The relationships between those 21keywords are shown in figure 4. The number of keyword co- occurrences can be determined by the size of the nodes in the image. The central nodes of Cluster I (red) contain the keywords: artificial intelligence, classification and cyber security, internet of things, machine learning, malware, malware detection and security. Cluster II (green) includes 7 items, anomaly detection, big data, cyber-security, deep learning and internet of things (IoT), intrusion detection and smart grid. Cluster III (blue) is mainly related to cyber security, data mining, feature selection, intrusion detection systems and network security. A time overlay visualisation in figure 5 demonstrates how in recent years the research has been shifted towards deep learning, machine learning, data mining and IoT.

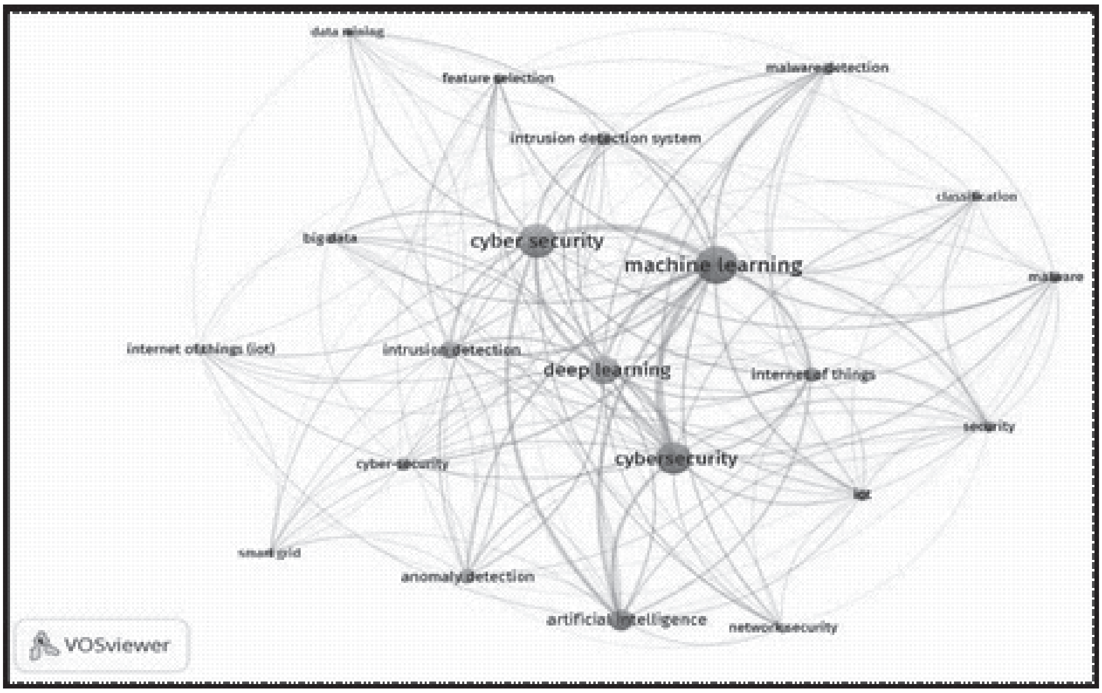

Figure 4: Co-Occurrences of author-keywords

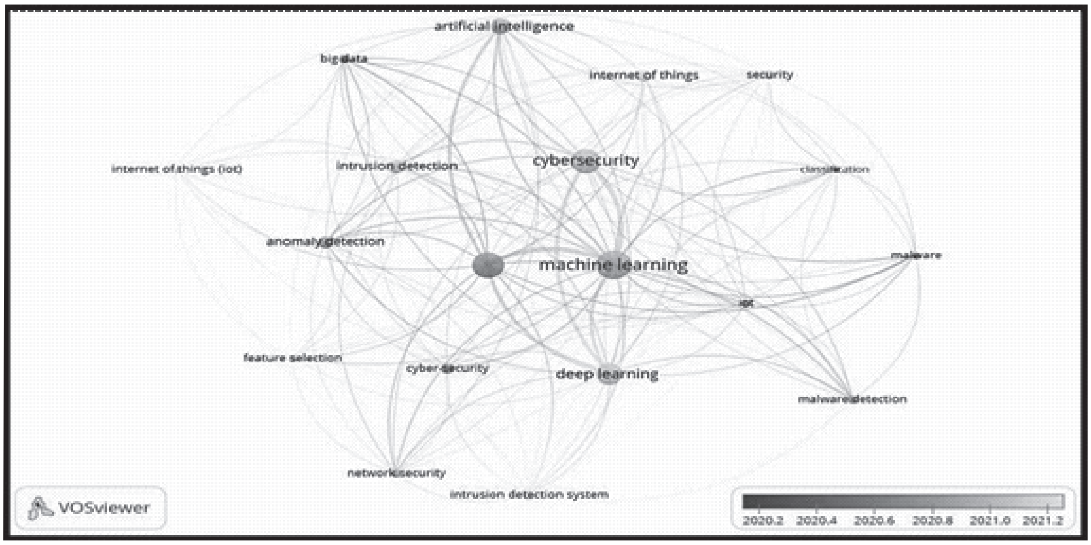

Figure 5: Co-occurrences of author-keywords (overlay visualisation)

8. Major findings

The following findings are drawn based on the analysis of the study conducted on 1065 articles in 391 sources, contributed by 3378 authors during the year 2013-2022.

- The growing trend in research is evident from the statistics. The fact that the number of publications has expanded exponentially over the last ten years suggests that the field is quite active and there is a steady increase in research interest.
- It reveals that IEEE Access, Sensors, Computers and Security are the key journals and got citations of 1472, 366 and 297 respectively. It recommended that researchers go through the aforementioned key journals to obtain a significant amount of information related to the fields.
- Within the countries, India ranks first.
- The Computer Science Department and Engineering are the main departments.
- ZHANG Jand LIU Y are the most productive authors from the year 2019
- Anomaly detection, artificial intelligence, big data, data mining, deep learning and internet of things are the keywords used by the authors
- The research has been shifted toward deep learning, machine learning, data mining and IoT in recent years

9. Conclusion

The paper's primary goal was to evaluate academic works on artificial intelligence based cyber security solutions published in the Scopus data base. The growing trend in research implies that the field is quite active and with the aid of AI, there is a steady increase in research interest in cyber security. The Computer Science Department and Engineering are the main departments conducting major research. It discloses that research into AI based cyber security is still in its early stages, and further advancement is anticipated in the years to come.

This study gives a useful overview of the current status of AI driven cyber security research. It highlights the literature progress rate, most effective authors, top publishing sources, top associated organisations and essential boundaries of the field. India has actively worked on strengthening its cyber security capabilities due to the increasing importance of digital technologies and the growing threat landscape. It is revealed from the study that multi-disciplinary evaluations of past and current works are required for future research in this area. In the future, researchers may under take comprehensive bibliometric analysis in specific areas of cyber security such as threat intelligence, data protection, security analytics, policy and governance among others.

References

Aria, M., & Cuccurullo, C. (2007). Bibliometrics: an R-tool for comprehensive science mapping analysis. *Journal of Informatics,* 11(4).

Bircan, T., & Salah, A. (2022). A bibliometric analysis of the use of artificial intelligence technologies for social sciences. *Mathematics,* 10(23), 4398. doi:10.3390/math10234398

Cheng, S., & Wang, B. (2012). An overview of publications on artificial intelligence research: a quantitative analysis on recent papers. *2012 Fifth International Joint Conference on Computational Sciences and Optimization.* doi:10.1109/cso.2012.156

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021a). How to conduct a bibliometric analysis: an overview and guidelines. *Journal of Business Research,* 133, 285-296.doi:10.1016/j.jbusres.2021.04.070

Dua, S., & Du, X. (2011). *Data mining and machine learning in cyber security.* Auerbach Publications.

Gupta, S. (2016). Scientometric mapping of research in library consortia. *International Journal of Digital Library Services,* 6(3).

Liu, S., You, S., Yin, H., Liu, S., Liu, Y., Yu, W., & Sundaresh, L. (2020). Model-free data authentication for cyber security in power systems. *IEEE Transactions on Smart Grid,* 11(5), 4565-4568.doi:10.1109/tsg.2020.2986704

Liu, Y., & Guo, Y. (2022). Towards real-time warning and defence strategy AI planning for cyber security systems aided by security ontology. *Electronics,* 11(24), 4128.doi:10.3390/electronics11244128

Luo, J., Hu, Y., & Bai, Y. (2021). Bibliometric analysis of the block chain scientific evolution: 2014-2020. *IEEE Access,* 9, 120227-120246 doi:10.1109/access.2021.3092192

Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed.). New York: Free Press of Glencoe.

Saif, A. N. M., & Purbasha, A. E. (2023). Cyberbullying among youth in developing countries: a qualitative systematic review with bibliometric analysis. *Children and Youth Services Review,* 146, 106831.doi: 0.1016/j.childyouth.2023.106831

Sharma, D., Mittal, R., Sekhar, R., Shah, P., & Renz, M. (2023). A bibliometric analysis of cyber security and cyber forensics research. *Results in Control and Optimization,* 10, 100204.doi:10.1016/j.rico.2023.100204

Talan, T. (2021). Artificial intelligence in education: a bibliometric study. *International Journal of Research in Education and Science,* 822-837.doi:10.46328/ijres.2409

Eck, N., & Waltman, L. (2018, April 27). *Vosviewer Manual.* Retrieved from https://www.vosviewer.com/

Yang, X., Shu, L., Liu, Y., Hancke, G. P., Ferrag, M. A., & Huang, K. (2022). Physical security and safety of IOT equipment: a survey of recent advances and opportunities. *IEEE Transactions on Industrial Informatics,* 18(7), 4319-4330.doi:10.1109/tii.2022.314 408