Chapter N 046

Bibliometric Analysis and Topic Modeling of the literature on Artificial Intelligence in Healthcare

**Abstract.** In the last five years there has been an accelerated growth in the scientific production on the subject of Artificial Intelligence and Healthcare by Scholars of the most diverse disciplines. Recently, the scientific corpus has been enriched with considerable literature reviews ranging from the overview of large collections of scientific documents to the recognition of the state of knowledge on specific aspects (e.g., in the medical field, ophthalmology, cardiology, nephrology, etc.). Following a bibliometric analysis of the literature on the subject, conducted on a vast collection of scientific contributions, we also searched for the "latent" themes in the semantic structures of these documents, identified the relationships between them and recognized those most likely to be investigated in the future. The methodological approach is located in the scientific field of relational bibliometry and Content Analysis. The results of the bibliometric analysis are presented in terms of: interactive maps of association of the contributions based on bibliographic coupling and subsequently on the co-occurrence of Author keywords.

**Keywords.** healthcare, artificial intelligence, bibliometric analysis, topic modeling.

# Introduction

The complex set of concepts and technologies which are generally referred to with the term "Artificial Intelligence" (AI) finds one of the most relevant and dynamic fields of application in healthcare. Artificial intelligence and health evoke two areas in impetuous evolution whose boundaries are widening with a frequency that does not seem risky to define daily. The combination of these two worlds, which at first glance seem to be far removed from each other, is increasingly closer and more natural. So much so that it has given rise to a new field called “Health AI” (Jung and Pfister, 2020; Martinho et al., 2021). In the spiral of reciprocal interdependencies between the growth in demand for collective and individual health - also driven by recent pandemic events - and the acceleration of the processes of digitization of socio-economic activities, the entire world of healthcare is in fact taking on new and rapidly evolving physiognomies; so much so that (Topol, 2019), in the subtitle of his book dedicated to Deep Medicine, introduces the reader to how “Artificial Intelligence can make healthcare human again”. The applications of computer systems capable of carrying out tasks that normally require human intelligence to health care activities are many and both the web and the scientific literature propose basic taxonomies, for some of which it is possible to refer to Wang et al. (2021), and Bohr and Memarzadeh (2020).

In recent years, in fact, the convergence of the evolution of AI with the increase in health data availability and processing power has given rise to a new class of AI solutions that are beginning to have a significant impact on all aspects of healthcare (Garg et al., 2021; Househ et al., 2021; Saxena et al., 2021). These solutions are based on advanced techniques of Machine Learning and Deep Learning, Natural Language Processing and Computer Vision able to process and interpret huge amounts of data (Big Data), all developed starting from the fifties, ideally in response to the famous question "can machines think?" of Turing (1950) and explored in some aspects in the following years [the ability of a machine to play checkers better than its programmer (Samuel, 1959), that of learning to make decisions (Quinlan, 1986) and many others]. In the the healthcare sectors big data originate from different sources (EHR - Electronic Health Records, wearable devices, patient-generated data, etc.) to support clinicians and decision-makers in various areas such as precision medicine, risk prediction and prevention, drug discovery and development, Robotic Process Automation in administrative and managerial tasks, etc. The healthcare sector is one of the most data-intensive sectors and, at the same time, one of the least digitized (Secinaro et al., 2021; Shen et al., 2021). In recent years, however, there has been a growing interest in the use of AI solutions also in healthcare, mainly driven by 3 factors:

* The exponential increase in data availability due to the growth of EHRs (Electronic Health Records) and other digital health initiatives;
* The rapid evolution of AI technologies, in particular Deep Learning;
* The need to address some of the main challenges facing the healthcare sector such as rising costs, chronic disease management and drug development.

Precision medicine is one of the most promising applications of AI in healthcare (Seyhan and Carini, 2019; Ahmed et al., 2020). AI can be used to analyze large amounts of data (genomic data, patient medical records, clinical trials data, etc.) to identify patterns and correlations that can help develop more personalized treatment plans for individual patients. In the past, precision medicine was limited by the lack of data and computing power. However, recent advances in AI and machine learning have made it possible to process large amounts of data quickly and accurately. As a result, precision medicine is now being used to develop more personalized treatment plans for individual patients.

There are many other potential applications of AI in healthcare (e.g., disease risk prediction, drug development, robotic process automation, clinical trials), and it is clear that the health sector is at the beginning of an AI-powered transformation that will have a profound impact on the way care is delivered and managed.

## Why another bibliometric review on AI and Healthcare Life?

In recent years, there has been an explosion of interest in bibliometric reviews. These reviews use quantitative methods to analyze the scholarly output of individuals, institutions, or entire fields of study. Bibliometric reviews can provide valuable insights into patterns of research productivity, emerging trends, and areas of potential impact. However, they are also time-consuming and expensive to produce. As a result, there is a need for careful justification when proposing a new bibliometric review. In this paper, we argue that there is still value to be gained from additional bibliometric reviews, particularly when they are designed to address specific research questions. We believe that well-designed bibliometric reviews can complement traditional narrative literature reviews and contribute to our understanding of the current state of scholarship in a given field.

A bibliometric review is a quantitative assessment of the scholarly literature in a given field. It is similar to a traditional literature review, but it goes beyond simply summarizing and critiquing the existing research. Bibliometric reviews also analyze the relationships between different pieces of research, identify trends and patterns, and assess the overall state of the field.

Despite their popularity, bibliometric reviews have been criticized for a variety of reasons. First and foremost, their approach is commonly challenged for being too basic and not taking into account significant elements such as publication bias. Second, their findings are frequently contested on the grounds that they are too broad and not precise enough to be useful to practitioners. Finally, the manner in which these evaluations are conducted is frequently attacked for being inefficient and time-consuming.

Despite these criticisms, bibliometric reviews can be a valuable tool for researchers and practitioners alike. When used correctly, they can provide insights into the latest trends and developments in a particular field of research. Additionally, they can be used to identify gaps in the literature and to inform future research directions. Overall, bibliometric reviews can be a useful addition to the literature on AI and Healthcare.

A "free" search on the Web conducted without claiming to be exhaustive yields numerous reviews of the literature on the topic under examination, including several dozen conducted with bibliometric methods. Among these, those complete with a searchable full-text, necessary to identify the databases and the software used, the research questions they answer, the limitations posed to the construction of the of the set of reference documents.

Table 1. Some of the existing bibliometric reviews.

Sources: Author’s elaboration

Our bibliometric review differs from others in two key ways. First, we cascade the bibliographic coupling and the co-occurrence of keywords (Iandolo et al., 2021). This allows us to identify not only which papers are citing each other, but also which keywords are being used in conjunction with each other. Second, we use a combination of manual and automated methods to identify the relationships between papers. This ensures that our results are both accurate and comprehensive.

# Material and methods

## Methods

In the early days of information science, people looked for ways to organize and make sense of the rapidly growing body of knowledge (De Solla Price, 1965). Among the methods conceived we find bibliographic coupling (Kessler, 1963), co-citation analysis (Marshakova, 1973; Small, 1973), the co-occurrence of keywords or words in abstracts and/or full text. All can be used to identify groups of papers on similar topics, as well as relationships between different fields of research. Bibliographic coupling of documents is a measure of the similarity between two documents, based on the number of references they share in common. Bibliographic coupling has been used to study a wide range of topics, from the history of science to the spread of ideas. It is a powerful tool for understanding the structure of knowledge, and it continues to be used by scholars all over the world. There are three main types of bibliographic coupling: content-based, co-citation, and network. Co-citation coupling measures the similarity between two documents based on the number of times they are cited by other documents. Keywords or words co-occurrence coupling measures the similarity between two documents based on the topics they discuss. Each of these methods has its own strengths and weaknesses, and all three can be used to complement each other in order to create a more complete picture of document similarity.

Our method consists in using bibliographic coupling in two successive phases in order to identify homogeneous networks of scientific papers and subsequently characterize these networks in terms of topics mainly dealt with through the co-occurrence of the authors' keywords.

## Tools

VOSviewer is a free, open-source software application for overlaying and visualizing sets of bibliographic data, such as journal article authors, co-authors, and citation networks. Developed by scholars at Leiden University in the Netherlands (van Eck and Waltman, 2010, 2014), VOSviewer has been designed to be used with a variety of different data sets and can be customized to suit the needs of specific research projects. In addition to its visualization features, VOSviewer also offers a number of analytical tools that can be used to examine relationships between different data sets. VOSviewer offers various options for customizing the visualizations. One option is to overlay the visualization with additional information. For example, users can overlay a map of the world onto their network visualization. This can be helpful for understanding the geographic relationships between nodes in the network. Other overlays include node labels, node sizes, and more. In addition to overlays, VOSviewer also offers options for changing the colors and shapes of nodes and edges. This allows users to create visualizations that are both visually appealing and informative.

# Data

"artificial intelligence" (Topic) and healthcare (Topic) and Proceedings Papers or Early Access or Book Reviews or Meeting Abstracts or Book Chapters or Editorial Materials or Letters or Data Papers or News Items or Corrections (Exclude – Document Types) and English (Languages) : 2164 documents, may 21.

Fig. 1. Documents by year and by source.

Sources: Authors’ elaboration based on data retrieved from Web of Science.

# Analysis

The result of the bibliographic coupling of the 2164 scientific publications making up our sample is summarized by the map in the figure.

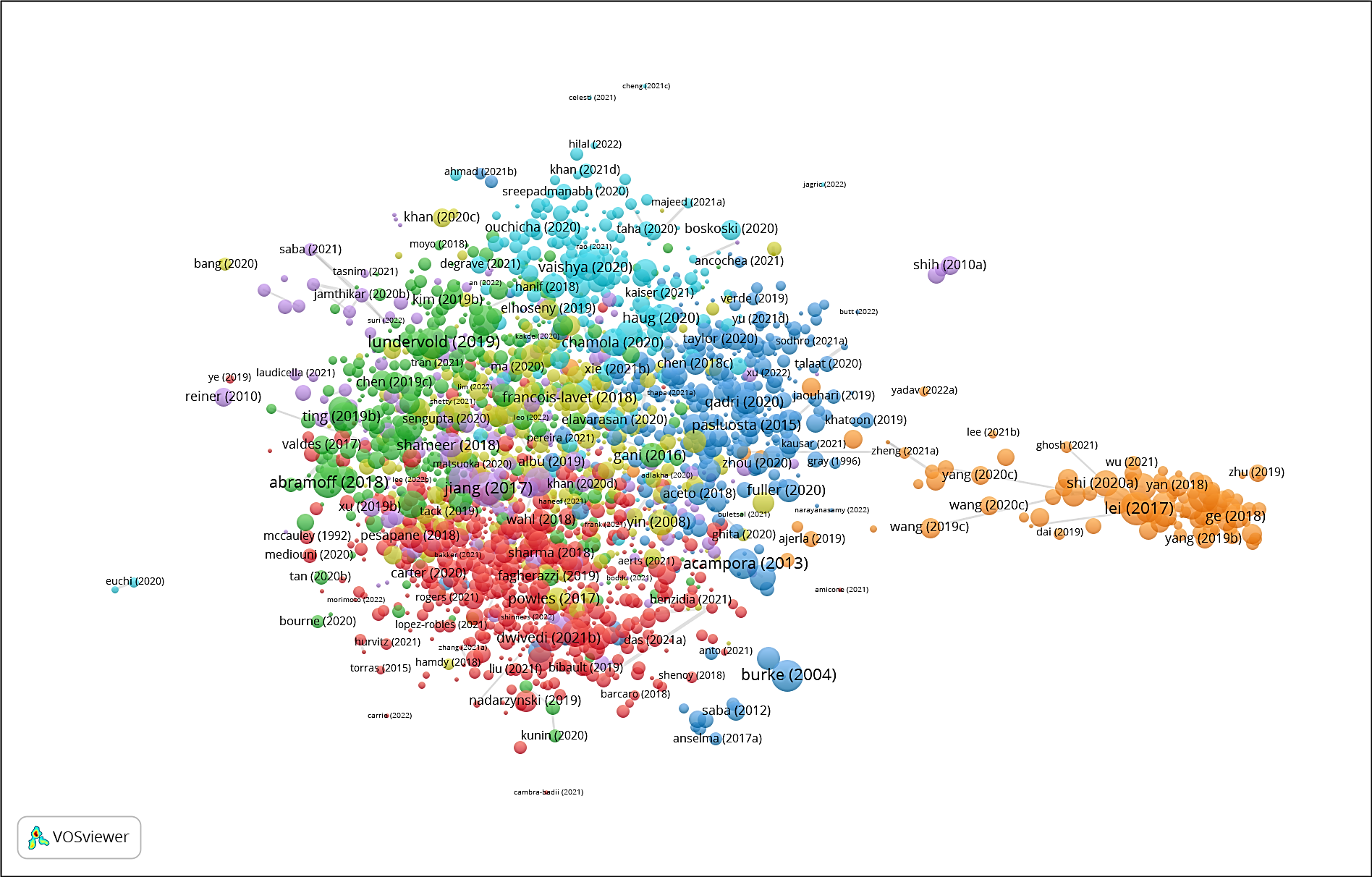
**

Fig. 3. Bibliographic coupling of the sample.

Sources: Authors’ elaboration.

The parameters used for the construction of the map were the following: Normalization method: Association strenght; Layout: Attraction 2, Repulsion 0; Clustering: Resolution 1.00, minimum cluster size 100. With these parameters, the clustering algorithm returns 7 clusters whose sizes are between 566 (contributions) of the first group and 103 of the last.

Cluster 1 (red): Ethics, Big Data, Digital Health. (Healthcare Sciences & Services; Medical Informatics; Computer Science) <https://bit.ly/3mbGkd5>

Cluster 2 (green): medical imaging, convolutional neural network/s, telemedicine (medicine, general & internal; computer science, information systems) https://bit.ly/3ap2dDa

Cluster 3 (blue): internet of things, blockchain, big data (computer science, information systems; engineering, electrical and electronic; telecommunications) <https://bit.ly/3x46hQM>

Cluster 4 (…): natural language processing, big data, explainable artificial intelligence (computer science, information systems; medical informatics) <https://bit.ly/3GJiOgQ>

Cluster 5 (…): big data, neural network/s, digital health (cardiac & cardiovascular; healthcare sciences & services) <https://bit.ly/3Md10vN>

Cluster 6 (…): covid-19, internet of things, convolutional neural network/s, diagnosis, big data <https://bit.ly/3x4Mi4y>

Cluster 7 (…): electronic skin, flexible electronics, wearable electronics, sensors, tactile sensor/s <https://bit.ly/3m9iHSm>

The characterization of the clusters in terms of topics mainly dealt with in each of them through the determination of the frequency of the keywords used by the authors required the preliminary identification of the contributions contained within each cluster and subsequent analysis in Vosviewer. The keywords of the emerging authors in each cluster are shown in the table.

The static representations of the bibliometric maps relating to the authors' keywords are omitted for the sake of brevity and as summarized in the table. In correspondence with each cluster, however, we have indicated the web page containing the interactive (zoomable and navigable) version of the relative map. All the maps are displayed in the "clean" version, ie cleaned of terms indicating the same concept (eg: "healthcare" and "healt care" or "artificial intelligence" and "ai").

Table 2. Most frequent keywords (topics most frequently dealt with) in each cluster.

Sources: Authors’ elaboration

# Conclusions

***Limitations and Implications for future research***

In addition to the natural limitations of bibliometric analysis methodologies (the dependence of the results on search strategies in bibliographic databases, meaning not only the formulation of the queries but also the series of possible limitations to the construction of the sample of bibliographic references to be analyzed, for example, but not limited to, the type of publication, the reference time extension, etc.) poorly highlighted in the existing reviews examined, this work highlights further limitations.

While this bibliometric review of the literature on artificial intelligence and healthcare allows to identify some privileged areas of attention by scholars of different disciplines (this is the case, for example, of ethics in the disciplinary field "healthcare science & services "Or" Internet of Things "in the computer science, information systems area) from another reveals the limits of hard clustering techniques, as demonstrated by the presence of some keywords in several groups (one for all, the keyword" big data ").

The numerous existing reviews (structured, bibliometric, systematic, etc.) must be integrated by reviews based on Topic Modeling techniques, which make it possible to identify topics, historical trends (classical and emerging topics), associations between the documents and to predict, on a probabilistic basis, which scientific fields will be most likely to see development in the future.

References

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. Q. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database-the Journal of Biological Databases and Curation, 35. doi: 10.1093/database/baaa010

Bohr, A., & Memarzadeh, K. (2020). Artificial Intelligence in Healthcare: Elsevier Science.

De Solla Price, D. J. (1965). Networks of Scientific Papers. Science, 149(3683), 510-515.

Garg, L., Basterrech, S., Banerjee, C., & Sharma, T. K. (2021). Artificial Intelligence in Healthcare: Springer Nature Singapore.

Gu, D. X., Li, K., Wang, X. Y., Li, X. G., Liu, F. H., Jiang, L., & Zhang, F. F. (2019). Discovering and Visualizing Knowledge Evolution of Chronic Disease Research Driven by Emerging Technologies. Ieee Access, 7, 72994-73003. doi: 10.1109/access.2019.2916630

Gu, D. X., Li, T. T., Wang, X. Y., Yang, X. J., & Yu, Z. R. (2019). Visualizing the intellectual structure and evolution of electronic health and telemedicine research. International Journal of Medical Informatics, 130, 11. doi: https://sci-hub.ru/10.1016/j.ijmedinf.2019.08.007

Gupta, A., & Katarya, R. (2020). Social media based surveillance systems for healthcare using machine learning: A systematic review. Journal of Biomedical Informatics, 108, 13. doi: https://sci-hub.ru/10.1016/j.jbi.2020.103500

Househ, M., Borycki, E., & Kushniruk, A. (2021). Multiple Perspectives on Artificial Intelligence in Healthcare: Opportunities and Challenges: Springer International Publishing.

Hussien, H. M., Yasin, S. M., Udzir, N. I., Ninggal, M. I. H., & Salman, S. (2021). Blockchain technology in the healthcare industry: Trends and opportunities. Journal of Industrial Information Integration, 22, 23. doi: https://sci-hub.ru/10.1016/j.jii.2021.100217

Iandolo, F., Vito, P., Loia, F., Fulco, I., & Calabrese, M. (2021). Drilling down the viable system theories in business, management and accounting: a bibliometric review. Systems Research and Behavioral Science, 38(6), 738-755.

Islam, M. M., Poly, T. N., Alsinglawi, B., Lin, L. F., Chien, S. C., Liu, J. C., & Jian, W. S. (2021). Application of Artificial Intelligence in COVID-19 Pandemic: Bibliometric Analysis. Healthcare, 9(4), 10. doi: https://sci-hub.ru/10.3390/healthcare9040441

Jung, H. H., & Pfister, F. M. J. (2020). Blockchain-enabled Clinical Study Consent Management. Technology Innovation Management Review, 10(2), 14-24. doi: 10.22215/timreview/1325

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. American documentation, 14(1), 10-25.

Marshakova, I. V. (1973). System of document connections based on references. Nauchno-tekhnicheskaya informatsiya seriya 2-informatsionnye protsessy I sistemy(6), 3-8.

Martinho, A., Kroesen, M., & Chorus, C. (2021). A healthy debate: Exploring the views of medical doctors on the ethics of artificial intelligence. Artificial Intelligence in Medicine, 121, 10. doi: 10.1016/j.artmed.2021.102190

Quinlan, J. R. (1986). Induction of Decision Trees. Machine Learning, 1(1), 81-106. doi: 10.1023/A:1022643204877

Saheb, T., Saheb, T., & Carpenter, D. O. (2021). Mapping research strands of ethics of artificial intelligence in healthcare: A bibliometric and content analysis. Computers in Biology and Medicine, 135, 19. doi: https://sci-hub.ru/10.1016/j.compbiomed.2021.104660

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development, 3(3), 210-229. doi: 10.1147/rd.33.0210

Saxena, A., Brault, N., & Rashid, S. (2021). Big Data and Artificial Intelligence for Healthcare Applications: CRC Press.

Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V., & Biancone, P. (2021). The role of artificial intelligence in healthcare: a structured literature review. Bmc Medical Informatics and Decision Making, 21(1), 23. doi: 10.1186/s12911-021-01488-9

Seyhan, A. A., & Carini, C. (2019). Are innovation and new technologies in precision medicine paving a new era in patients centric care? Journal of Translational Medicine, 17, 28. doi: 10.1186/s12967-019-1864-9

Shen, L., Bai, J. W., Wang, J., & Shen, B. R. (2021). The fourth scientific discovery paradigm for precision medicine and healthcare: Challenges ahead. Precision Clinical Medicine, 4(2), 80-84. doi: 10.1093/pcmedi/pbab007

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(4), 265-269.

Sohn, E., Noh, K. R., Lee, B., & Kwon, O. J. (2018, Aug 28-31). Bibliometric Network Analysis and Visualization of Research and Development Trends in Precision Medicine. Paper presented at the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Barcelona, SPAIN.

Topol, E. (2019). Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again: Basic Books.

Turing, A. M. (1950). I.—Computing Machinery and Intelligence. Mind, LIX(236), 433-460. doi: 10.1093/mind/LIX.236.433

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538. doi: 10.1007/s11192-009-0146-3

Van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau & D. Wolfram (Eds.), Measuring scholarly impact (pp. 285-320): Springer.

Wang, Q., Su, M., Zhang, M., & Li, R. R. (2021). Integrating Digital Technologies and Public Health to Fight Covid-19 Pandemic: Key Technologies, Applications, Challenges and Outlook of Digital Healthcare. International Journal of Environmental Research and Public Health, 18(11), 50. doi: https://sci-hub.ru/10.3390/ijerph18116053