Preprints.org

Review

Analyzing the Scholarly Footprint of ChatGPT: Mapping the Progress and Identifying Future Trends

Faiza Farhat , Emmanuel Sirimal Silva , Hossein Hassani , Dag Øivind Madsen [*] , Shahab Saquib Sohail , Yassine Himeur , M. Afshar Alam , Aasim Zafar

Review

Analyzing the Scholarly Footprint of ChatGPT: Mapping the Progress and Identifying Future Trends

Faiza Farhat [1], Emmanuel Sirimal Silva [2], Hossein Hassani [3], Dag Øivind Madsen [4,*], Shahab Saquib Sohail [5], Yassine Himeur [6], M. Afshar Alam [5] and Aasim Zafar [7]

[1] Department of Zoology, Aligarh Muslim University, Aligarh, India

[2] Department of Economics and Law, Glasgow School for Business and Society, Glasgow Caledonian University, Glasgow, United Kingdom

[3] The Research Institute of Energy Management and Planning (RIEMP), University of Tehran, Tehran, Iran

[4] USN School of Business, University of South-Eastern Norway, Hønefoss, Norway

[5] Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard, New Delhi, India

[6] College of Engineering and Information Technology, University of Dubai, Dubai, UAE

[7] Department of Computer Science, Aligarh Muslim University, Aligarh, India

* Correspondence: dag.oivind.madsen@usn.no

Abstract: This paper presents a comprehensive analysis of the scholarly footprint of ChatGPT, an AI language model, using bibliometric and scientometric methods. The study aims to understand the evolution of research output, citation patterns, collaborative networks, application domains, and future research directions related to ChatGPT. By analyzing data from the Scopus database, 533 relevant articles were identified for analysis. The findings reveal the prominent publication venues, influential authors, and countries contributing to ChatGPT research. Collaborative networks among researchers and institutions are visualized, highlighting patterns of co-authorship. The application domains of ChatGPT, such as customer support and content generation, are examined. Moreover, the study identifies emerging keywords and potential research areas for future exploration. The methodology employed includes data extraction, bibliometric analysis using various indicators, and visualization techniques such as Sankey diagrams. The analysis provides valuable insights into ChatGPT's influence in academia and offers researchers guidance for further advancements. This study stimulates discussions, collaborations and innovations to enhance ChatGPT's capabilities and impact across domains.

Keywords: ChatGPT; bibliometric analysis; scientometric methods; research trends; citation analysis; collaborative networks; application domains; future directions

1. Introduction

The rapid advancements in artificial intelligence (AI) have led to the development of sophisticated language models that can understand and generate human-like text. One such notable AI language model is ChatGPT (https://openai.com/chatgpt), an autoregressive language model that uses deep learning techniques to generate coherent and contextually relevant responses to user inputs. Since its launch, ChatGPT has gained significant attention and adoption in various domains, including content generation, healthcare, education, data science, tourism, and customer support/assistance [1–7]. The introduction of ChatGPT has sparked discussions and debates surrounding its potential implications in various domains [see for example, 8,9,10]. Notably, issues

related to ethical considerations and biases [6], and the impact of large language models on knowledge assessment [11,12] have garnered attention in recent discourse. Scholarly investigation of ChatGPT has emerged as a critical area of research, aiming to understand its impact, applications, and future directions [3,6,13].

To date, however, only a few authors have used bibliometric and scientometric methods to analyze ChatGPT. Khosravi, et al. [14] carried out an analysis of the broader chatbot literature, while Levin, et al. [15] used bibliometrics to explore publications on ChatGPT in the field of obstetrics and gynecology. Our study differs by focusing especially on ChatGPT publications, and by considering the latest developments up until June 2023. In our view, bibliometric and scientometric analysis can provide valuable insights into the research landscape surrounding ChatGPT, including the evolution of research output, citation patterns, collaborative networks, application domains, and emerging research trends. By analyzing a comprehensive dataset of scholarly publications, this study aims to shed light on the scholarly footprint of ChatGPT and its influence in academia.

This study employs a multifaceted approach, utilizing bibliometric and scientometric methods to analyze the scholarly footprint of ChatGPT. Bibliometric analysis [see for example, 16] offers a quantitative approach to evaluate the scholarly impact of ChatGPT research. By carefully gathering and analyzing relevant data from the Scopus database, we address several pivotal research questions:

➢ Publication Trends: How has research output related to ChatGPT evolved over time? Which are the prominent publication venues and journals that feature research on ChatGPT?

➢ Citation Analysis: How has ChatGPT been referenced in scholarly literature? Which papers, authors, countries and journals have made significant contributions to the understanding and advancement of ChatGPT?

➢ Collaborative Networks: Who are the key contributors and collaborators in the ChatGPT research landscape? What patterns of collaboration and co-authorship exist among researchers, institutions and countries working on ChatGPT-related topics?

➢ Application Domains: In which primary domains has ChatGPT found an application? How are researchers leveraging its capabilities in fields such as customer support, content generation, and virtual assistance?

➢ Future Directions: Based on the keyword analysis related to ChatGPT's scholarly footprint, what emerging keywords and potential research area for future research can be identified? What challenges and opportunities lie ahead in enhancing ChatGPT's capabilities and impact?

By addressing these research questions, this paper aims to provide a comprehensive and up-to-date analysis of ChatGPT's scholarly footprint. Our findings not only contribute to a better understanding of ChatGPT's influence in academia but also serve as valuable insights for researchers interested in the development and utilization of AI language models. Ultimately, through mapping its progress and identifying future trends, we hope to stimulate discussions, collaborations, and innovations that drive the continued advancement of ChatGPT and its applications across various domains.

By utilizing data from the Scopus database, we identified 533 relevant articles published between November 2022 and May 2023 that focus on ChatGPT. The selected articles underwent thorough evaluation based on various criteria, including organization, country/region, journal, total citations, and keywords. This analysis revealed several key insights as presented in our findings and discussion later on. For example, there has been a remarkable surge in scholarly publications related to ChatGPT, with 533 articles produced within a short span of six months. This indicates a thriving research interest and highlights the growing recognition of the potential applications of ChatGPT. Furthermore, the high collaboration rate of 88.91% among authors suggests a strong community of researchers working on ChatGPT, sharing ideas and resources to advance the field. In addition, we also uncover interesting details around the publication venues contributing to ChatGPT research

which evidences its impact in diverse scientific disciplines, the contributions of different countries to ChatGPT research, top authors and institutions.

The remainder of this paper is organized such that Section 2 presents the methodology, Section 3 provides an overview of the main findings. Section 4 discusses the findings in relation to the existing literature on ChatGPT. Finally, Section 5 concludes the paper by highlighting contributions, limitations and ideas for further works.

2. Methodology

2.1. Data Extraction

Figure 1 provides an overview of the methodology. On June 6, 2023, we used the Scopus database to search for articles that contain the search queries "chatgpt", "ChatGPT", or "Chat-GPT" in the title, abstract, or keywords. We selected 533 articles for our bibliometric review study after applying exclusion and inclusion criteria. We evaluated the retrieved articles using the following criteria: organization, country/region, journal, total citations, and keywords. We downloaded the complete records for bibliometric analysis and imported them into the Biblioshiny (Bibiliometrix) and VOSviewer software packages. Various indicators have been used in the literature for bibliometric analysis, including total article count, average citations per article (ACPA), total citation count, total link strength, and Hirsch index (H-index). These metrics are commonly used in bibliometric studies, with the H-index being a widely recognized measure of research quality and quantity for authors and research areas [17]. ACPA is also widely accepted as a measure of research impact for individual works, authors, and publication outlets. Citation analysis is conducted to explore the scientific impact and themes of the study under consideration, and co-authorship and co-occurrence have also been investigated to analyze scientific collaboration. Three-field Sankey diagrams are also used to identify the relationship among three interrelated sets of values [18]. All of these indicators have been taken into account in this bibliometric study.
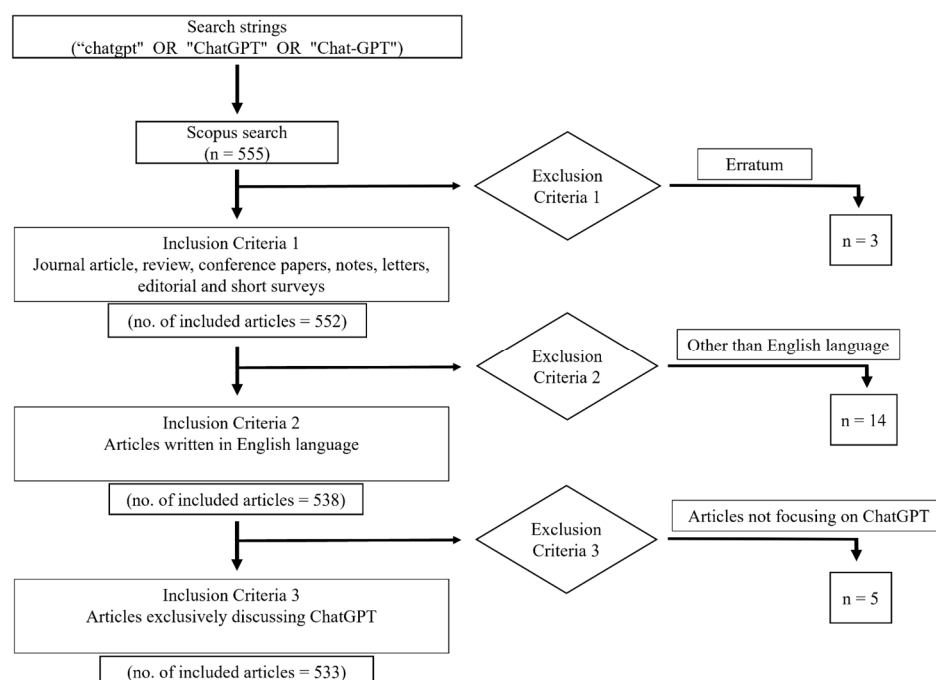
Figure 1. Methodology.

2.2. Data Analysis

A comprehensive analysis of ChatGPT research was conducted, encompassing 533 publications from 87 countries and 1195 institutions. These publications, originating from 341 different sources,

were authored by 1434 individuals and received a total of 1362 citations. Moreover, a total of 1998 keywords were identified. The analysis involved employing the full counting approach, which focuses on elements connected to one another. This approach facilitated citation analysis and co-authorship analysis. Collaboration networks among authors, institutions, and countries were visualized using illustration maps. The size of the circles in these maps indicated the strength and frequency of collaborations between individuals and organizations.

Additionally, citation maps displayed the connections and citations between different partners, with larger circles representing higher citation counts and stronger linkages. To analyze the relationships between keywords, a keyword map was generated using the complete counting method. To examine the interactions among three distinct interconnected variables, three-field Sankey diagrams were utilized. These diagrams enable the analysis of relationships involving authors, author's keywords, and keywords. Similarly, the interplay between country, publication source, and keywords, as well as author, title-term, and source, were also investigated using these diagrams. Furthermore, the research trend and popular topics in ChatGPT research were explored through the identification of significant research terms, word cloud analysis, and examination of keyword co-occurrence. This map grouped related keywords together and assigned equal weight to each co-occurrence link. Consequently, terms with higher frequency were represented by larger circles in the map.

3. Findings

We begin by providing a comprehensive overview of the research conducted on ChatGPT during the period of 2022-2023. In a short span of only 6 months (Nov. 2022 to May 2023), a total of 533 documents were produced from 341 sources, from 87 different countries involving 1434 authors, indicating a thriving research interest (Table 1). The annual growth rate of 17,566.67% signifies a remarkable surge in scholarly publications related to ChatGPT and the growing interest in the potential applications of ChatGPT. The total corpus involved 1434 authors, with 159 of them contributing to single-authored documents. This represents a significant collaboration rate of 88.91%, highlighting the collaborative network of the research area. This collaboration suggests that there is a strong community of researchers working on ChatGPT, and that they are sharing ideas and resources to advance the field. Among the documents, 420 were single-country contributions, while 113 demonstrated collaboration between multiple countries. The involvement of 1195 institutions highlighted diverse organizational contributions.

Table 1. Overview of the retrieved data related to ChatGPT.

Description	Results
Timespan	2022-2023 (Nov. 2022 to May 2023)
Sources (Journals, Books, etc.)	341
Documents	533
Annual Growth Rate %	17566.67
Total citations	1362
Self-Citations	824
Self-Citations %	60.5
Average citations per doc	2.546
References	11244
DOCUMENT CONTENTS	
Total Keywords	1998
Keywords Plus (ID)	1371
Author's Keywords (DE)	882
AUTHORS	
Authors	1434
Authors of single-authored docs	159

Single-authored docs	182
Single-authored docs %	34.14
Co-Authors per Doc	3.08
Authors collaboration %	88.91
COUNTRIES	
Countries	87
Single-country docs	420
Multiple-country docs	113
Countries collaboration %	21.2
INSTITUTIONS	
Institutions	1195
Institutions collaboration %	6.44

3.1. Types of Documents Published and the Thematic Area of Research

Out of the total of 533 publications obtained, a significant portion comprises empirical papers, representing 36.77% (196 articles) of the corpus. Letters constitute 19.51% (104 articles), editorials make up 18.57% (99 publications), while notes account for 14.55% (Figure 2). Interestingly, it is observed that besides empirical papers, a substantial portion of the ChatGPT corpus consists of letters, notes, and editorials, making up 52.63% of the total publications. On the other hand, the number of review articles published is relatively low.
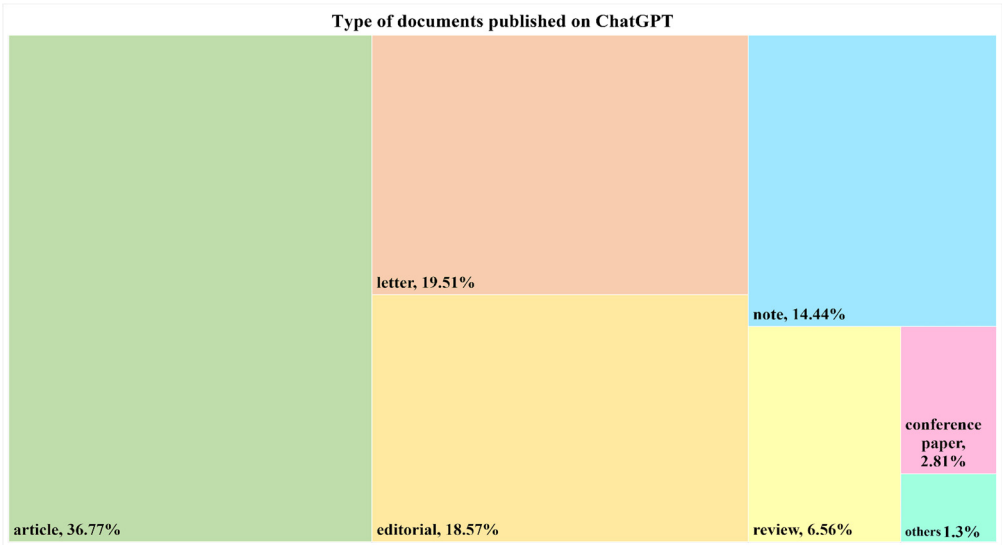

Figure 2. Tree map representing the type of documents published on ChatGPT.

Table 2. Top 10 most relevant journals based on article count.

Journal	Article Count	Citation Count	Average Citation per Article (ACPA)	H-Index	Publisher
Annals Of Biomedical Engineering	28	46	1.64	150	Springer Netherlands
Nature	17	367	21.58	1331	Nature Publishing Group
Library Hi Tech News	12	20	1.66	22	Emerald Group Publishing Ltd.
Medical Teacher	6	3	0.5	131	Informa Healthcare
Radiology	6	100	16.66	320	Radiological Society of North America Inc.
Accountability In Research	5	11	2.2	35	Taylor and Francis Ltd.

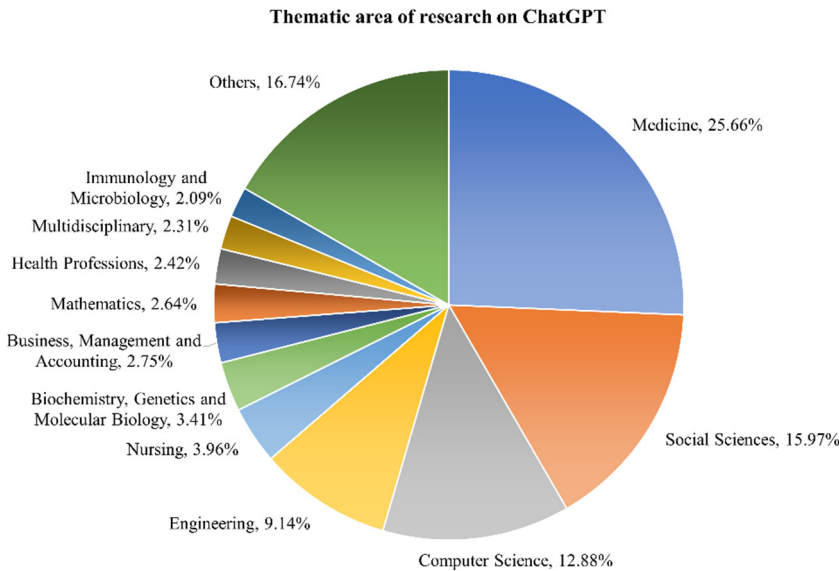|					
Annals Of Surgical Oncology	5	1	0.2	192	Springer New York
IEEE/CAA Journal of Automatica Sinica	5	19	3.8	67	IEEE Advancing Technology for Humanity
JMIR Medical Education	5	39	7.8	23	JMIR Publications Inc.
Journal of Chemical Education	5	0	0	95	American Chemical Society

Figure 3. Thematic subject categories of research on ChatGPT.

There are 533 publications relating to ChatGPT in 341 different journals. The top 10 journals account for 17.63% of the corpus and 44.65% of the total citations. Annals of Biomedical Engineering has thus far published the most articles (28), followed by Nature (17) and Library Hi Tech News (12). Nature was the most cited journal (367 citations), followed by Radiology (100) and Science (93). Based on their H-index, Nature ranks first (1331), Radiology ranks second (320) and Annals of Surgical Oncology ranks third (192).

A total of 87 different countries have contributed to research on the topic of ChatGPT. Table 3 displays the top 10 countries on the basis of article count. The USA ranked first in terms of publications, with 173 articles, accounting for over 32.54% of the entire corpus. India (48 articles, 9%) and UK (47 articles, 8.81%) rank second and third, respectively, in terms of contribution. In addition, the USA has a more considerable global academic impact than any other country, as demonstrated by the highest citation count (391). The UK ranks second in terms of citations with 153. Moreover, countries such as Australia, China, and Italy have made significant contributions, with citation counts of 76, 73, and 67, respectively.

Table 3. Top 10 most relevant countries based on article count.

Country	Article Count	Total Citations	Average Citation per Article (ACPA)
United States	173	391	2.26
India	48	50	1.04
United Kingdom	47	153	3.25
China	43	73	1.69
Australia	38	76	2
Canada	23	25	1.08
Italy	23	67	2.91
Germany	21	59	2.80

South Korea	15	33	2.2
France	14	57	4.07

Table 4 presents the top 10 authors and their corresponding article metrics. Wang F.Y. from the Institute of Automation Chinese Academy of Sciences in China stands out with the highest article count of 9 and a total of 23 citations. Following closely is Wu H. from Duke University School of Medicine in the USA, with 7 articles and 16 citations. Interestingly, Kleebayoon A. from Joseph Ayo Babalola University in Nigeria and Wiwanitkit V. from Chandigarh University in India have published 6 articles each but have not received any citations. On the other hand, authors with a lower article count have also garnered significant citation numbers. For instance, Ali M.J. from L.V. Prasad Eye Institute in India, Lu Y. from Zhengzhou University in China, and Gu S. from Duke University School of Medicine in the USA have achieved 8, 10, and 15 citations, respectively, with just 5 articles each. Among the top 10 authors, three authors represent China and three represent the USA, while two hail from India.

Table 4. Top 10 most relevant authors based on article count.

Authors	Article Count	Total Citations	Average Citation per Article	Affiliation	Country of Origin
Wang F.Y.	9	23	2.55	Institute of Automation Chinese Academy of Sciences	China
Wu H.	7	16	2.28	Duke University School of Medicine	USA
Cheng K.	6	12	2	Zhengzhou University	China
He Y.	6	16	2.66	The University of North Carolina at Chapel Hill	USA
Kleebayoon A.	6	0	0	Joesph Ayobabalola University	Nigeria
Teixeira Da Silva J.A.	6	5	0.83	Miki-cho Post Office, Kagawa	Japan
Wiwanitkit V.	6	0	0	Chandigarh University	India
Ali M.J.	5	8	1.6	L.V. Prasad Eye Institute India	India
Gu S.	5	15	3	Duke University School of Medicine	USA
Lu Y.	5	10	2	Zhengzhou University	China

A total of 1195 institutions have contributed to the 533 publications, with Duke University participating in the most papers (14). Chinese Academy of Sciences (9), Chandigarh University (8) make up the top 3 organizations based on article count (Table 5). Duke University has received the most citations, cited 32 times, followed by Chinese Academy of Sciences and University of Chinese Academy of Sciences with 23 citations each. In terms of average citations per article, University of Chinese Academy of Sciences takes the top position with 3.28, followed by the Beijing Sport University with 2.66. Among the top 10 institutions, 5 institutions are from China representing the highest contribution to the field and 2 from the USA.

Table 5. Top 10 most relevant Institutions based on article count.

Organization	Article Count	Total Citations	Average Citation per Article	Country of Origin
Duke University	14	32	2.28	USA
Chinese Academy of Sciences	9	23	2.55	China
Chandigarh University	8	02	0.25	India
Johns Hopkins School of Medicine	7	08	1.14	USA
Tianjin Medical University	7	16	2.28	China
University of Chinese Academy of Sciences	7	23	3.28	China
Beijing Sport University	6	16	2.66	China
University of Toronto	6	03	0.5	Canada
Zhengzhou University	6	12	2	China
Monash University	6	06	1	Australia

3.2. Most Cited Documents, Authors, Countries and Journals

When the citation network analysis was carried out in VOSviewer, it was observed that 34 articles have at least 10 citations, 15 articles have 20 citations and only 5 articles have received 50 citations (Figure 4a). The size of the circle denotes the number of citations, and the connecting lines represent their citation network. Larger the circle larger would be the citation count of an article and more connecting lines reflect that the articles are citing another article or cited by other articles [19]. A total of 22 articles organized in 8 different clusters are linked among each other with 28 links (Figure 4b). The largest citation network is associated with Sallam [7] with 13 links, followed by Biswas [20] and Dwivedi, Kshetri, Hughes, Slade, Jeyaraj, Kar, Baabdullah, Koohang, Raghavan and Ahuja [4] with 5 links independently. The most cited document in the field of ChatGPT research is the editorial titled "ChatGPT is fun, but not an author" by Thorp [21], with 93 citations. The second most cited document is a note titled "ChatGPT listed as author on research papers: many scientists disapprove" by Stokel-Walker [22], with 88 citations. These two influential works have played a significant role in raising awareness and initiating critical conversations around the ethical implications of attributing authorship to AI language models.

(a)

(b)

(c)

(d)

10
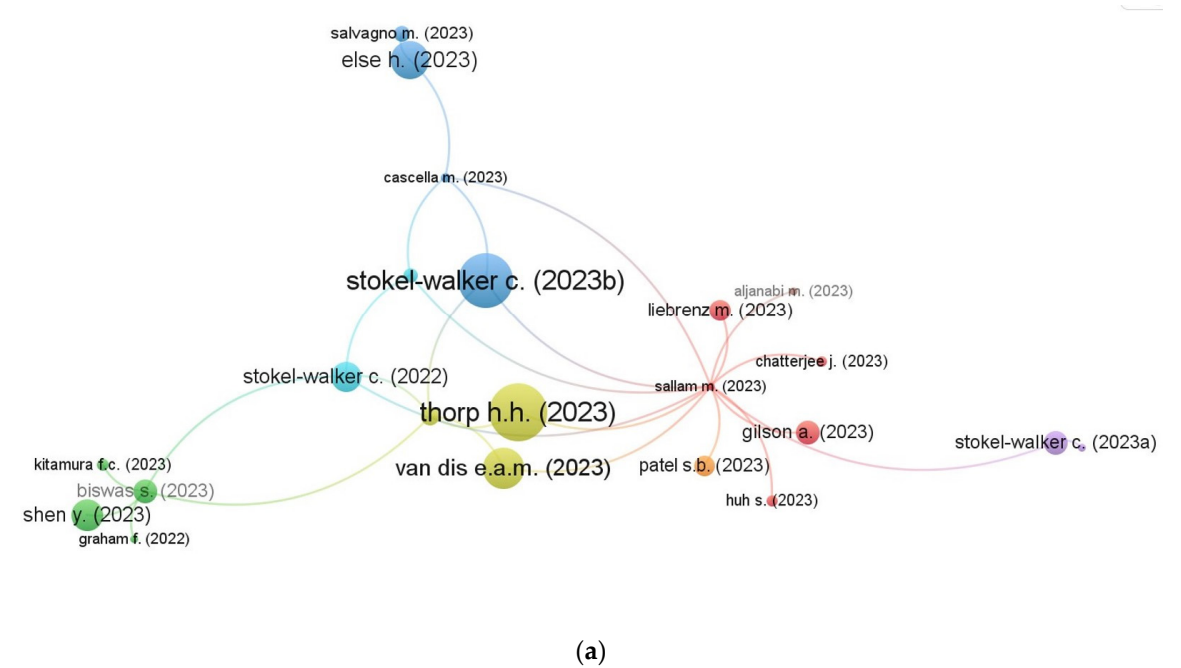
Figure 4. (a): Most cited articles and their citation network. (b):Most cited authors and their citation network. (c); Most cited countries and their citation network. (d): Most cited journals and their citation network.

The top 20 most cited documents are listed in Table 6. Of the top 20 most cited documents, 8 are notes, 6 are editorials, 4 are articles, and only 2 are reviews. The only research article that is among the top 10 most cited documents is "How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment" [12], otherwise it is either a note or editorial which forms the top 10 most cited document list. 7 of the top 20 most cited documents have been published by Nature, while Radiology and The Lancet Digital Health each have published 3.

Table 6. Top 20 most cited articles on ChatGPT.

No.	Title	Total Citation	Article Type	Journal	Country of First Author	Reference
1	ChatGPT is fun, but not an author	93	Editorial	Nature	USA	[21]
2	ChatGPT listed as author on research papers: many scientists disapprove	88	Note	Nature	UK	[22]
3	ChatGPT: five priorities for research	63	Note	Nature	Netherlands	[23]
4	Tools such as ChatGPT threaten transparent science; here are our ground rules for their us	61	Editorial	Nature	USA	[24]
5	Abstracts written by ChatGPT fool scientists	57	Note	Nature	USA	[25]
6	ChatGPT and Other Large Language Models Are Double-edged Swords	47	Editorial	Radiology	USA	[26]
7	AI bot ChatGPT writes smart essays — should professors worry?	43	Note	Nature	UK	[27]
8	ChatGPT and the Future of Medical Writing	33	Note	Radiology	USA	[20]
9	What ChatGPT and generative AI mean for science	33	Note	Nature	UK	[28]
10	How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment	33	Article	JMIR Medical Education	USA	[12]
11	Generating scholarly content with ChatGPT: ethical challenges for medical publishing	29	Note	The Lancet Digital Health	Switzerland	[29]
12	ChatGPT: the future of discharge summaries?	28	Note	The Lancet Digital Health	UK	[30]
13	Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education	28	Article	Journalism and Mass Communication Educator	USA	[31]
14	"So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and	24	Article	International Journal of	India	[4]

	implications of generative conversational AI for research, practice and policy			Information Management		
15	Can artificial intelligence help for scientific writing?	21	Article	Critical Care	Belgium	[32]
16	ChatGPT: evolution or revolution?	19	Editorial	Medicine, Health Care and Philosophy	Ireland	[33]
17	ChatGPT: friend or foe?	18	Editorial	The Lancet Digital Health		[34]
18	Chatting about ChatGPT: how may AI and GPT impact academia and libraries?	16	Review	Library Hi Tech News	USA	[35]
19	ChatGPT Is Shaping the Future of Medical Writing But Still Requires Human Judgment	15	Editorial	Radiology	Brazil	[36]
20	What Does ChatGPT Say: The DAO from Algorithmic Intelligence to Linguistic Intelligence	15	Review	IEEE/CAA Journal of Automatica Sinica	China	[37]

The citation analysis of authors visualizes the most cited authors and their citation network. It is observed that 187 authors have at least 10 citations. Stokel-Walker, C. is the most cited author, followed by Thorp, H.H., Bockting, C.L. and Else, H. with 93, 63 and 57 citations, in respective order (Figure 4b). Besides having the most citations, Stokel-Walker, C. has the largest citation network with 93 citing partners. The second largest citation network is associated with Biswas, S. with 88 different citing partners. The most frequent citing partners are Wu, H. and Cheng, K., cited each other at least 13 times, the next in line are Wu, H. and Lu, Y. having 12 link strength.

The citation analysis of countries showed that a large number of countries are actively citing each other's work. There are 38 countries that have received at least 10 citations, 29 countries that have received at least 20 citations, and 10 countries that have received at least 50 citations. The citation network of countries is very dense, meaning that there are a lot of connections between countries (Figure 4c). The most citing partners are the United States and the United Kingdom, with a link strength of 31. The United States and India are the second most citing partners, with a link strength of 30. The United States and Australia are the third most citing partners, with a link strength of 25. In terms of citation network, the United States has the largest network, with 35 links. India and China are tied for second place, with 30 links each.

The citation network analysis of journals revealed that the largest citation network consisted of 22 journals citing each other frequently (Figure 4d). 25 journals have at least 10 citations. The most cited journal on the topic of ChatGPT is Nature, with 367 citations, followed by Radiology with 100 citations and Science with 93 citations. Nature and Annals of Biomedical Engineering are the most frequent citing partners (Link strength 6). Afterwards, Nature along with Healthcare, Radiology and Library Hitech News makes the next frequent citing partners citing each other at least 4 times.

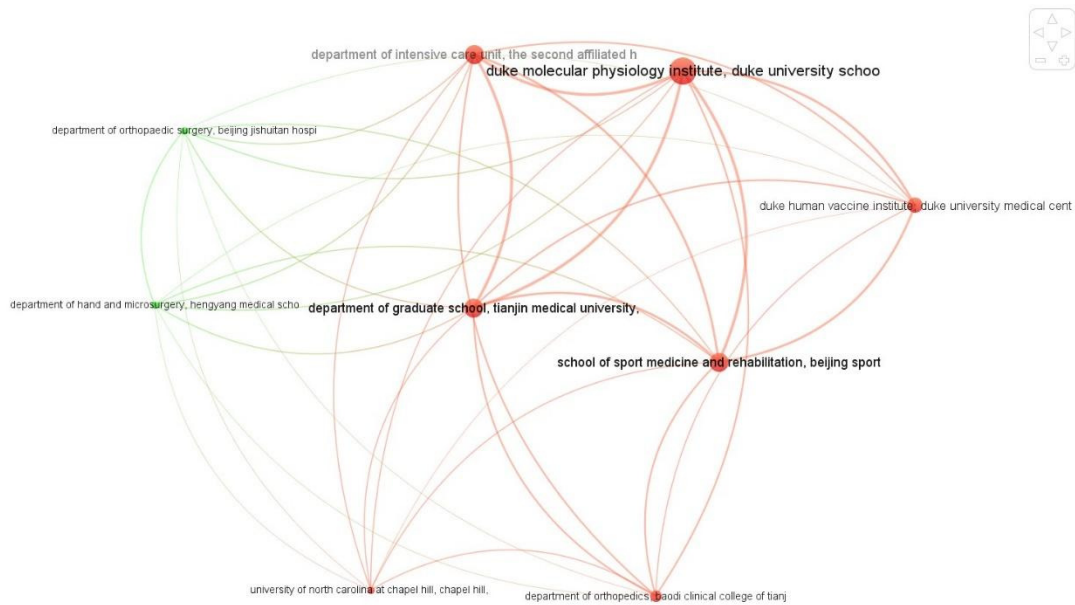### 3.3. Collaboration Network of Author, Institution and Countries

Of the 118 authors who have published at least two articles on ChatGPT, only 29 have collaborated with each other. These 29 authors are divided into five clusters, with the largest cluster (cluster 1) consisting of 10 authors. The second largest cluster (cluster 2) consists of 8 authors, followed by cluster 3 (5 authors), cluster 4 (4 authors), and cluster 5 (2 authors) (Figure 5). The two most collaborative authors, Wang, F.Y. and Wang, X, belong to the cluster 1 with 11 (17 link strengths) and 9 (14 link strengths) collaborations, respectively. Afterwards, Li, Z. with a link strength of 13 and 9 collaboration contribute to the 5th cluster. All the 8 authors of green cluster Wu, H., Quo, Q., Hey, Y., Lu, Y., Gu, S, Cheng, K, Li, C and Xie, R., have 8 collaborations each. Wu, H with Cheng, K. and

Hey, Y. are the most frequent collaborating partners (Link strength). Wu, H. and Gu, S. are the second most collaborative partners with a link strength of 5.
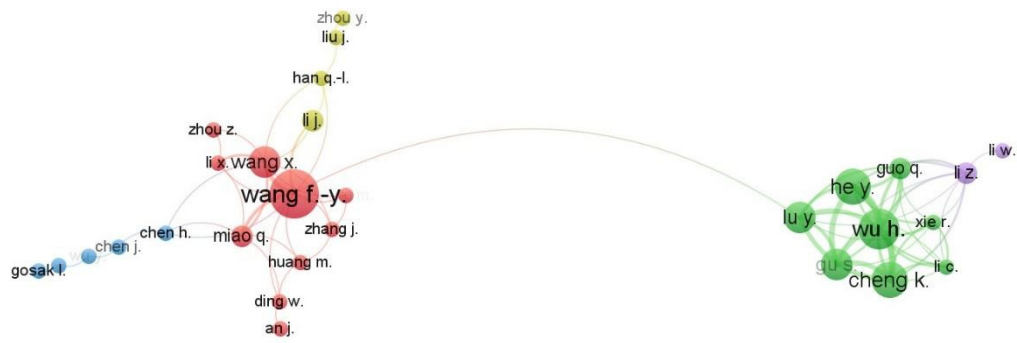
Figure 5. Largest collaboration network of authors.

Out of 1195 institutions around the world, 53 have published at least two articles on ChatGPT. The largest collaborating network consists of only nine institutions, six from China and three from the United States (Figure 6). All nine institutions have an equal number of collaborating links, with eight each. However, *Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, NC, United States* has published the most articles on ChatGPT in collaboration with eight different institutions. It is followed by *Department of Graduate School, Tianjin Medical University, Tianjin, China, Department of Intensive Care Unit, The Second Affiliated Hospital of Zhengzhou University, Henan, Zhengzhou, China,* and *School of Sport Medicine and Rehabilitation, Beijing Sport University, Beijing, China,* which have published five articles each.

Figure 6. Largest collaboration network of institutions.

A total of 56 countries have published at least two papers on ChatGPT, and 53 countries have collaborated with each other on these papers. The United States has the most collaborating partners (35 countries), with China being the most frequent collaborator (link strength of 12) (Figure 7). The

United Kingdom, India, Australia, and Italy are also frequent collaborators (link strength of 7 each). Australia and the United Kingdom have the second most collaborating partners (29 countries each). Afterwards, India and Nigeria, and India and Cambodia are the most frequent collaborators (link strength of 6 each). The United States also leads in single-country publications (SCP) with 75 articles, followed by China and India with 25 and 22 articles, respectively (Figure 8). China has the highest number of multiple-country publications (MCP) with 12 articles, followed by the United States with 9 articles. The United Kingdom and Italy have 7 and 6 MCP, respectively.
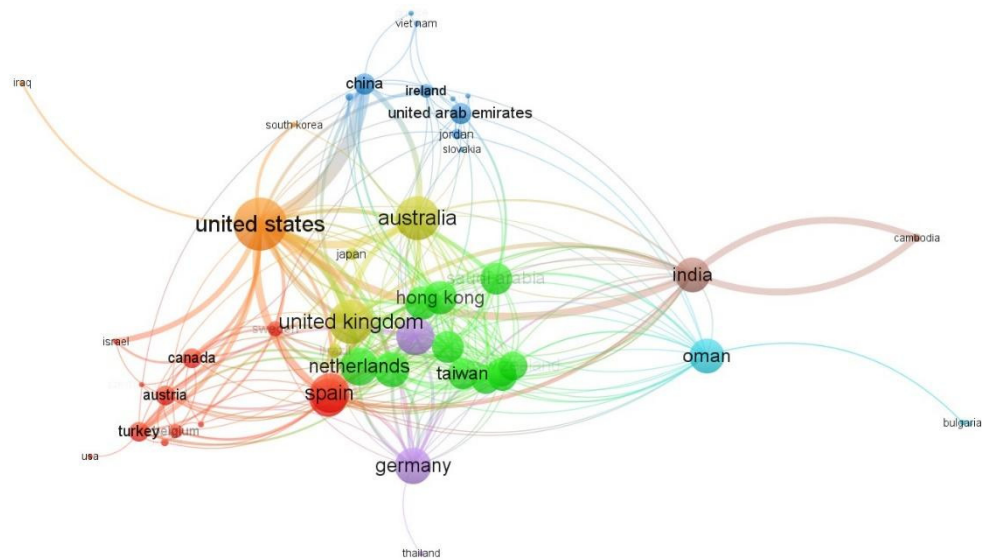

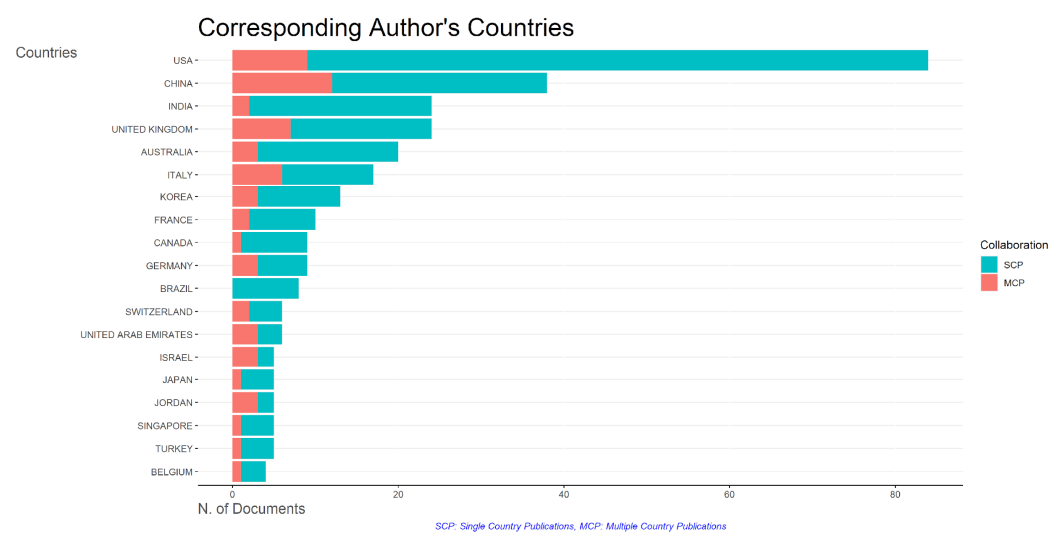**Figure 7. Largest collaboration network of countries.**

Figure 8. Graph representing single country and multiple country publications on ChatGPT.

3.4. Sankey Diagram (Three-Field Plot)

A Sankey diagram, also known as a three-field plot, is used to visualize the flow of values from one set to another. Figure 9 illustrates the relationship among the author's country, sources of their publications, and keywords chosen by them. The analysis reveals that authors from the United States have published their articles in 12 different journals, indicating a wide range of publication sources compared to other countries. In contrast, Chinese authors predominantly publish their works in three journals: Annals of Biomedical Engineering, IEEE Transaction on Vehicle, and IEEE/CAA Journal of

Automatica Sinica. Canadian and Australian authors exhibit the next highest levels of publication diversity, with 7 and 6 different journals, respectively.

Annals of Biomedical Engineering is a favored choice among authors from seven different countries, with the majority of publications coming from China, the United States, and India, in respective order. The most commonly selected keywords by authors include "ChatGPT," "artificial intelligence," "natural language processing," "large language model," "chatbot," and "machine learning." Notably, the most diverse keyword is "ChatGPT," followed by "artificial intelligence," which is highly popular among authors as well as sources. Among the journals, Library Hi Tech News has indexed 13 out of the top 20 most frequently used author's keywords whereas Nature have only three keywords in common with author's keywords viz., "machine learning", "ethics" and "education".
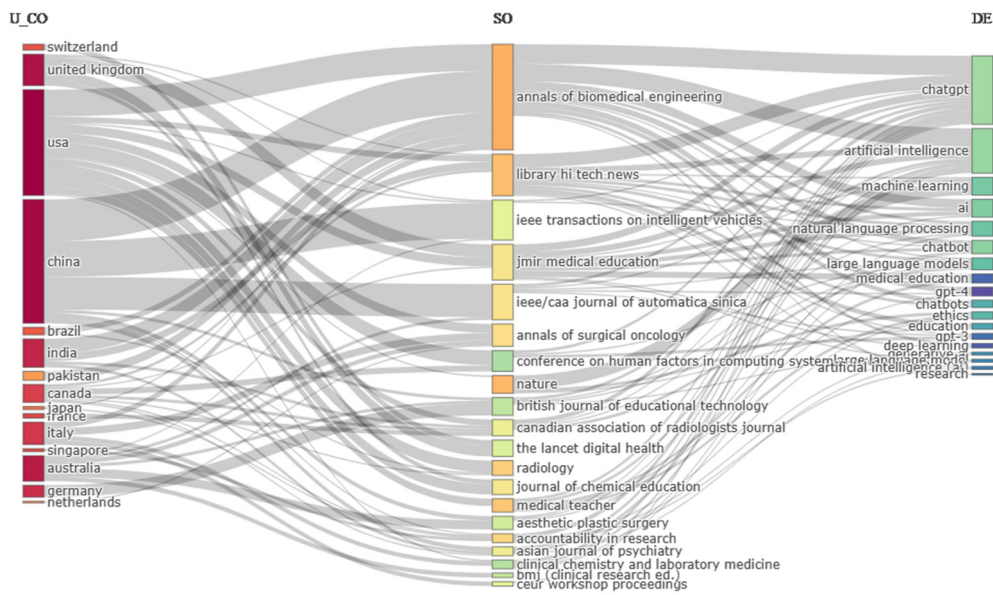
Figure 9. Three-field plot of countries, journals and Author's keywords. AU_CO: Author's countries, SO: source, and DE: Author's keywords.

Figure 10 shows the relationship between author keywords, authors, and keywords plus. Author keywords are chosen by authors, while keywords plus are automatically chosen by journals based on the frequency of cited and referenced title words. It is observed that author keywords and keywords plus are quite different from each other. For example, "ChatGPT" is the most frequently used keyword by authors, while keywords plus tend to favor "artificial intelligence." There are some common keywords in both categories, but their frequencies vary. For example, "ChatGPT" is a favorite choice of authors, but it is one of the least appearing keywords plus. Notably, authors such as Wu, H., Cheng, K., Hey, Y., Gu, S., and Lu, Y., share common keywords that fall under both keyword's categories, viz., "artificial intelligence," "chatbot," and "chatbots."
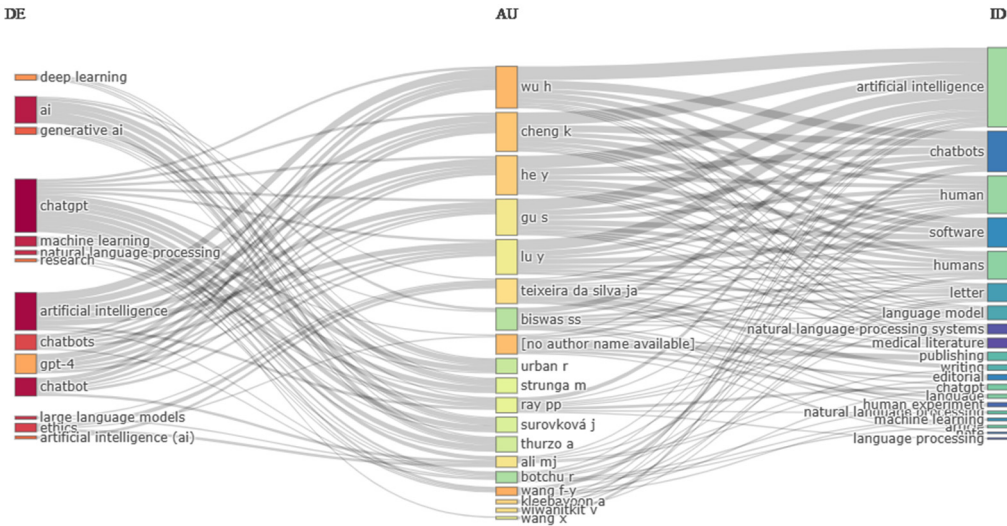
Figure 10. Three-field plot of keyword plus, authors and author's keywords. DE: Author's keywords, AU: Author and ID: Keyword plus.

The three-field plot Sankey diagram (Figure 11) shows the relationship between author, title-term used by them and sources. It is obvious that "ChatGPT" is the most widely used title-term by the authors as well as the most widely accepted title-term of journal publications. Terms like "Intelligence" and "potential" in the titles of publication shows the trending research topics related to the ChatGPT. Apart from machine learning related terms such as "AI", "language" "model", "artificial" and "intelligence", most frequent title-terms are "medical", "academic", "writing", "education", "medicine" etc. reflecting the recent thrust area of ChatGPT research. These title-terms are very frequently accepted by the top journals like Nature, Annals of Biomedical Engineering, Radiology and Library Hi Tech News.
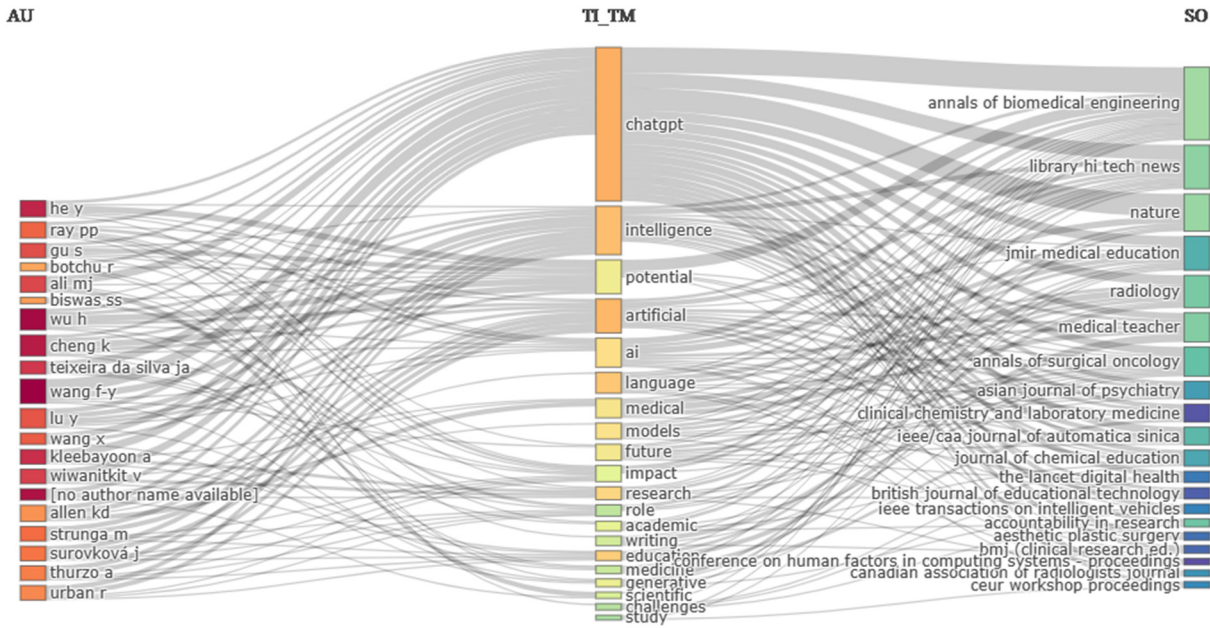

Figure 11. Three-field plot of authors, title-terms and sources. AU: Author and Tl_TM: Title-term and SO: Source.

3.5. Keyword Analysis

For keyword analysis, the most relevant keywords are retrieved using Bibliometric software package. As Figure 12 shows, the most occurred keyword is "Artificial intelligence" with 205 occurrences, followed by "human", "humans", "language" and "chatgpt" with 151, 94, 55 and 39 occurrences, respectively. Other keywords with high occurrences are "article", "natural language processing", "publishing" and "writing". Additionally, a word cloud of the most frequent keywords is plotted to illustrate the highly used terms in the field of ChatGPT research (Figure 13).

Figure 12. Most relevant keywords and their occurrences.

Figure 13. Word cloud of keywords on ChatGPT.

To conduct co-occurrence analysis of keywords, a threshold of at least 10 occurrences was chosen, resulting in a selection of 49 keywords that appeared at least 10 times (Figure 5). Synonyms of the keywords were excluded, and network maps were generated to visualize the top 5 most frequently occurring keywords and their co-occurring keywords. It was observed that "artificial intelligence" (Figure 15K1) and "chatgpt" (Figure 15K2) were the two most commonly co-occurring keywords in ChatGPT literature, each appearing alongside 47 distinct keywords. The third and fourth most frequently co-occurring keywords were "human" (Figure 15K3) and "natural language processing" (Figure 15K4), with co-occurrence network strengths of 46 and 43, respectively. "Machine learning" secured the fifth position, co-occurring with 39 different keywords in the network (Figure 15K5).

Furthermore, Table 7 presents the top 10 pairs of keywords with the highest frequency of co-occurrence. The pair "artificial intelligence" and "chatgpt" exhibited the most frequent co-occurrence, appearing together 117 times. The second most frequent pair was "artificial intelligence" and "human," which co-occurred 102 times. Additionally, "human" and "chatgpt" were found to co-occur 38 times.
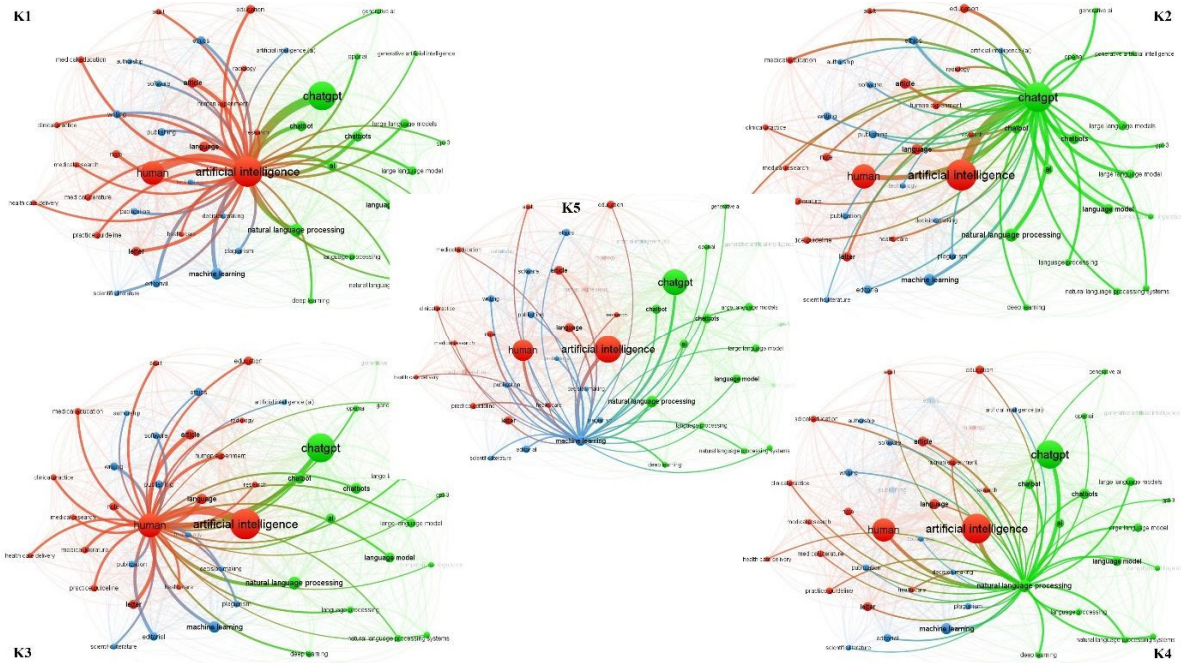

Figure 15. Top 5 most occurred keywords and co-occurred keyword their network.

Table 7. Top 10 pairs of co-occurred keywords.

No.	Keyword 1	Keyword 2	Co-Occurrence
1	Artificial intelligence	ChatGPT	117
2	Artificial intelligence	Human	102
3	Human	ChatGPT	38
4	Artificial intelligence	Natural language processing	33
5	Human	Article	32
6	Artificial intelligence	Machine learning	29
7	Artificial intelligence	Article	28
8	ChatGPT	Natural language processing	27
9	Artificial Intelligence	Chatbot	27
10	ChatGPT	Chatbot	27

4. Discussion

The comprehensive analysis of ChatGPT research conducted in the period from November 2022 to May 2023 reveals a thriving research interest in the field. During this short timeframe, a total of 533 documents were produced, indicating a significant surge in scholarly publications related to ChatGPT. The annual growth rate of 17,566.67% highlights the growing interest in the potential applications of ChatGPT and the varied and significant impact it has had on humans across the globe. Google Trends for the search term 'ChatGPT' during this same period shows how the interest in the topic increased gradually over time reaching its peak in April 2023 and maintained a stable level of interest since.
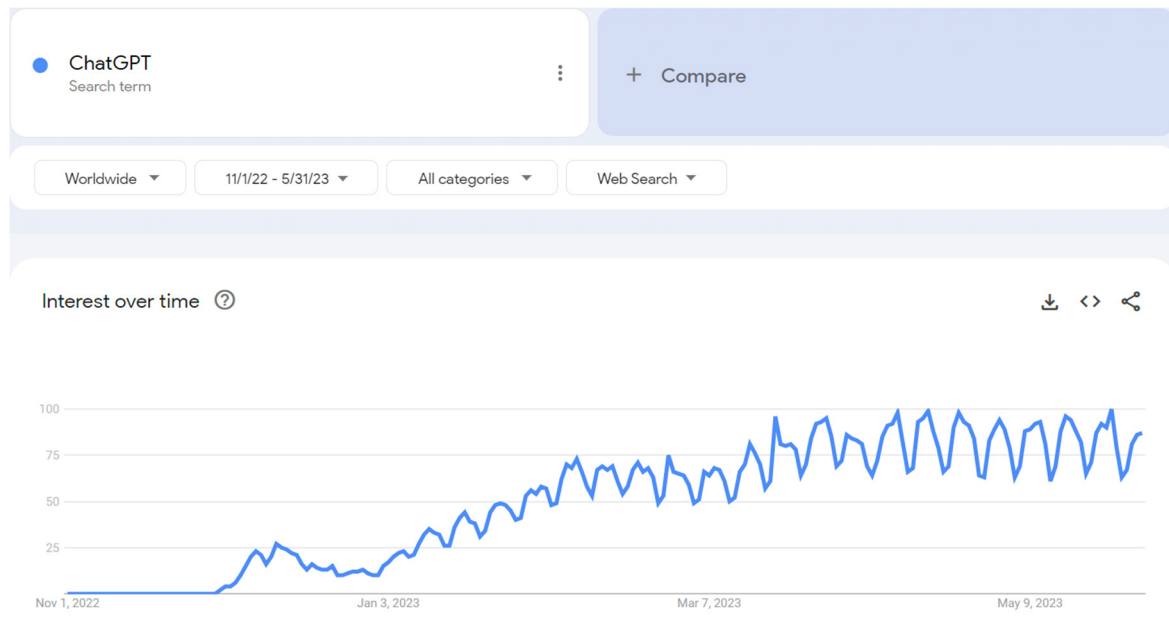
Figure 16. Worldwide Google Trends for 'ChatGPT' from 1st November 2022 - 31st May 2023.

Collaboration among researchers is evident in the analysis, with a high collaboration rate of 88.91% observed among the authors. This suggests a strong community of researchers working on ChatGPT who are actively sharing ideas and resources to advance the field. The involvement of 1195 institutions from various countries further emphasizes the collaborative nature of the research. This finding also appears to indicate two other aspects. First, the power of technological innovations and artificial intelligence in drawing researchers from varied backgrounds to collaborate and share knowledge leading to inter and multidisciplinary outputs. Secondly, the desire to be the first to publish and establish one's footprint in this emerging and disruptive field, whereby collaboration enables a speedy and meaningful route to swift publication.

The type of documents published on ChatGPT shows a diverse range of contributions. Empirical papers constitute the largest portion of the documents, followed by letters, editorials, and notes. The significant presence of letters, notes, and editorials within the corpus indicates that there is a variety of perspectives and opinions surrounding ChatGPT. This also highlights the newsworthiness and hype underlying the emergence of ChatGPT. Furthermore, these types of outputs are also a route to swift publication which benefits authors not only in terms of enabling themselves to gain recognition as key thinkers within the field but also benefit from a potential surge in citations. Researchers should take caution when citing and referring to work that is published in outlets which are not peer-reviewed as the information contained could be misleading in some cases leading to a flawed impression of this emerging field. We also uncovered that the number of review articles published is relatively low, suggesting an area for further exploration and synthesis of existing knowledge.

The analysis of the top journals reveals the leading platforms for ChatGPT research. Annals of Biomedical Engineering published the highest number of articles, followed by Nature and Library Hi Tech News. Nature also stands out as the most cited journal, indicating its influence and reputation in the field. Given the wide-ranging implications of ChatGPT we would expect the list of journals that feature relevant research to expand exponentially over the coming months and years as the understanding of the implications of this innovation improves over time. In terms of countries, the United States emerges as the most prolific contributor with the highest number of publications. India and the United Kingdom follow closely behind. The USA also demonstrates the highest citation count, indicating its global academic impact. Other countries such as Australia, China, and Italy have also made significant contributions to ChatGPT research.

The top authors in the field showcase their contributions and impact. Wang F.Y. from the Institute of Automation Chinese Academy of Sciences leads with the highest article count, while

authors from Duke University School of Medicine in the USA also feature prominently. Notably, authors with a lower article count have achieved significant citation numbers, highlighting the quality and impact of their work. The top institutions contributing to ChatGPT research represent a mix of organizations from different countries. Duke University participates in the most papers, followed by Chinese Academy of Sciences and Chandigarh University. Duke University also received the highest number of citations, indicating the institution's research excellence and impact. The analysis of citation networks reveals the most cited documents, authors, countries, and journals. "ChatGPT is fun, but not an author" by Thorp [21] emerges as the most cited document, followed by "ChatGPT listed as author on research papers: many scientists disapprove" by Stokel-Walker [22]. These documents highlight the discussions and controversies surrounding ChatGPT and its use in research. The presence of notes and editorials in the top cited documents suggests that discussions and opinions are driving the conversation in the field.

Keyword analysis is an essential aspect of bibliometric research, providing insights into the most relevant terms and their co-occurrence patterns [38]. In the case of ChatGPT, a bibliometric software package was utilized to retrieve the most occurred keywords. Among these keywords, "Artificial intelligence" emerged as the most frequent, appearing 205 times. Following closely were "human," "humans," "language," and "ChatGPT " with 151, 94, 55, and 39 occurrences, respectively. The emergence of 'human' and 'humans' as significant co-occurring keywords is important given the nature of the innovation. In a world where artificial intelligence is taking over and automating many processes there is considerable concern on its impact on human nature which could result in significant political and economic issues if not addressed and considered carefully. For example, even before the emergence of ChatGPT, Hassani, et al. [39] argued on the importance of focusing on intelligence augmentation as the way forward and the urgent need for ethical frameworks that can regulate the growth of AI whilst protecting the wellbeing and interest of humans. Other significant keywords included "article," "natural language processing," "publishing," and "writing."

5. Conclusion

In this paper, we have carried out a comprehensive bibliometric analysis of the scholarly footprint of ChatGPT, which has shed light on the progress and future trends in the field of AI language models. By employing bibliometric and scientometric methods, we have explored various dimensions of ChatGPT research, including overall publication trends, citation patterns, collaborative networks, application domains, and future directions.

The analysis of publication trends revealed a remarkable surge in scholarly output related to ChatGPT within a short time frame of about six months. The analysis also examines the publication venues contributing to ChatGPT research and evidences the impact of ChatGPT on diverse scientific disciplines. Furthermore, the study explores the contributions of different countries to ChatGPT research and finds that the United States has the most significant global academic impact in the field of ChatGPT, but other countries such as China, Australia, and Italy have also made notable contributions to ChatGPT research. In terms of influential authors, Wang F.Y. from the Chinese Academy of Sciences and Wu H. from Duke University are among the top authors based on article count and total citations.

This study serves as a valuable resource for researchers, offering a comprehensive understanding of the scholarly footprint of ChatGPT. It can serve as a quick guide for new researchers who orient+ themselves in the landscape of GPT research, by highlighting the most influential authors, studies, and institutions thus far. Moreover, the findings can guide future research endeavors, collaborations, and innovations in enhancing ChatGPT's capabilities and impact. By mapping the progress and identifying future trends, we aim to stimulate discussions and contribute to the continuous advancement of ChatGPT and its applications across domains.

However, since the research literature on ChatGPT is rapidly expanding, there is much uncertainty related to its future evolution and trajectory. Donthu, Kumar, Mukherjee, Pandey and Lim [16, p. 295] point out that "bibliometric studies can only offer a short-term forecast of the research field" and highlight that it is important to be careful when making assertions about its future

importance and impact. Therefore, it becomes of great importance to revisit and update the findings of this bibliometric study. Such analyses can also shed light on the dynamics and evolution of the scientific field and community involved in the field of ChatGPT research.

References

1. Carvalho, I.; Ivanov, S. ChatGPT for tourism: applications, benefits and risks. *Tourism Review* **2023**, *ahead-of-print*, doi:10.1108/TR-02-2023-0088.
2. Hassani, H.; Silva, E.S. The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. *Big Data and Cognitive Computing* **2023**, *7*, 62.
3. Sohail, S.S.; Farhat, F.; Himeur, Y.; Nadeem, M.; Madsen, D.Ø.; Singh, Y.; Atalla, S.; Mansoor, W. The Future of GPT: A Taxonomy of Existing ChatGPT Research, Current Challenges, and Possible Future Directions. *Current Challenges, and Possible Future Directions (April 8, 2023)* **2023**.
4. Dwivedi, Y.K.; Kshetri, N.; Hughes, L.; Slade, E.L.; Jeyaraj, A.; Kar, A.K.; Baabdullah, A.M.; Koohang, A.; Raghavan, V.; Ahuja, M. "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management* **2023**, *71*, 102642.
5. Wood, D.A.; Achhpilia, M.P.; Adams, M.T.; Aghazadeh, S.; Akinyele, K.; Akpan, M.; Allee, K.D.; Allen, A.M.; Almer, E.D.; Ames, D. The ChatGPT Artificial Intelligence Chatbot: How Well Does It Answer Accounting Assessment Questions? *Issues in Accounting Education* **2023**, 1-28.
6. Ray, P.P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems* **2023**, *3*, 121-154, doi:https://doi.org/10.1016/j.iotcps.2023.04.003.
7. Sallam, M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. In Proceedings of the Healthcare, 2023; p. 887.
8. Lo, C.K. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences* **2023**, *13*, 410.
9. Ivanov, S.; Soliman, M. Game of algorithms: ChatGPT implications for the future of tourism education and research. *Journal of Tourism Futures* **2023**, *ahead-of-print*, doi:10.1108/JTF-02-2023-0038.
10. Baumgartner, C. The potential impact of ChatGPT in clinical and translational medicine. *Clinical and translational medicine* **2023**, *13*.
11. Farhat, F.; Sohail, S.S.; Madsen, D.Ø. How trustworthy is ChatGPT? The case of bibliometric analyses. *Cogent Engineering* **2023**, *10*, 2222988, doi:10.1080/23311916.2023.2222988.
12. Gilson, A.; Safranek, C.W.; Huang, T.; Socrates, V.; Chi, L.; Taylor, R.A.; Chartash, D. How does ChatGPT perform on the United States medical licensing examination? The implications of large language models for medical education and knowledge assessment. *JMIR Medical Education* **2023**, *9*, e45312.
13. Dave, T.; Athaluri, S.A.; Singh, S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Frontiers in Artificial Intelligence* **2023**, *6*, 1169595.
14. Khosravi, H.; Shafie, M.R.; Hajiabadi, M.; Raihan, A.S.; Ahmed, I. Chatbots and ChatGPT: A Bibliometric Analysis and Systematic Review of Publications in Web of Science and Scopus Databases. *arXiv preprint arXiv:2304.05436* **2023**.
15. Levin, G.; Brezinov, Y.; Meyer, R. Exploring the use of ChatGPT in OBGYN: a bibliometric analysis of the first ChatGPT-related publications. *Archives of Gynecology and Obstetrics* **2023**, doi:10.1007/s00404-023-07081-x.
16. Donthu, N.; Kumar, S.; Mukherjee, D.; Pandey, N.; Lim, W.M. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research* **2021**, *133*, 285-296, doi:https://doi.org/10.1016/j.jbusres.2021.04.070.
17. Farhat, F.; Athar, M.T.; Ahmad, S.; Madsen, D.Ø.; Sohail, S.S. Antimicrobial resistance and machine learning: past, present, and future. *Frontiers in Microbiology* **2023**, *14*, 1717.
18. Aria, M.; Cuccurullo, C. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics* **2017**, *11*, 959-975, doi:https://doi.org/10.1016/j.joi.2017.08.007.
19. Van Eck, N.J.; Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* **2010**, *84*, 523-538.
20. Biswas, S. ChatGPT and the future of medical writing. **2023**, *307*, e223312.
21. Thorp, H.H. ChatGPT is fun, but not an author. *Science* **2023**, *379*, 313, doi:10.1126/science.adg7879.
22. Stokel-Walker, C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature* **2023**, *613*, 620-621, doi:10.1038/d41586-023-00107-z.
23. van Dis, E.A.M.; Bollen, J.; Zuidema, W.; van Rooij, R.; Bockting, C.L. ChatGPT: five priorities for research. *Nature* **2023**, *614*, 224-226, doi:10.1038/d41586-023-00288-7.
24. Editorial. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* **2023**, *613*, 612, doi:10.1038/d41586-023-00191-1.

25.   Else, H. Abstracts written by ChatGPT fool scientists. *Nature* **2023**, *613*, 423, doi:10.1038/d41586-023-00056-7.

26.   Shen, Y.; Heacock, L.; Elias, J.; Hentel, K.D.; Reig, B.; Shih, G.; Moy, L. ChatGPT and other large language models are double-edged swords. **2023**, *307*, e230163.

27.   Stokel-Walker, C. AI bot ChatGPT writes smart essays — should academics worry? *Nature* **2022**, doi:10.1038/d41586-022-04397-7.

28.   Stokel-Walker, C.; Van Noorden, R. What ChatGPT and generative AI mean for science. *Nature* **2023**, *614*, 214-216, doi:10.1038/d41586-023-00340-6.

29.   Liebrenz, M.; Schleifer, R.; Buadze, A.; Bhugra, D.; Smith, A. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. *The Lancet Digital Health* **2023**, *5*, e105-e106.

30.   Patel, S.B.; Lam, K. ChatGPT: the future of discharge summaries? *The Lancet Digital Health* **2023**, *5*, e107-e108.

31.   Pavlik, J.V. Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism and Mass Communication Educator* **2023**, doi:10.1177/10776958221149577.

32.   Salvagno, M.; Taccone, F.S.; Gerli, A.G. Can artificial intelligence help for scientific writing? *Critical care* **2023**, *27*, 1-5.

33.   Gordijn, B.; Have, H. ChatGPT: evolution or revolution? *Medicine, Health Care and Philosophy* **2023**, doi:10.1007/s11019-023-10136-0.

34.   The Lancet Digital, H. ChatGPT: friend or foe? *The Lancet Digital Health* **2023**, *5*, e102, doi:10.1016/S2589-7500(23)00023-7.

35.   Lund, B.D.; Wang, T. Chatting about ChatGPT: how may AI and GPT impact academia and libraries? *Library Hi Tech News* **2023**, *40*, 26-29.

36.   Kitamura, F.C. ChatGPT is shaping the future of medical writing but still requires human judgment. **2023**, *307*, e230171.

37.   Wang, F.-Y.; Miao, Q.; Li, X.; Wang, X.; Lin, Y. What does ChatGPT say: The DAO from algorithmic intelligence to linguistic intelligence. *IEEE/CAA Journal of Automatica Sinica* **2023**, *10*, 575-579.

38.   Farhat, F.; Sohail, S.S.; Siddiqui, F.; Irshad, R.R.; Madsen, D.Ø. Curcumin in Wound Healing—A Bibliometric Analysis. *Life* **2023**, *13*, 143.

39.   Hassani, H.; Silva, E.S.; Unger, S.; TajMazinani, M.; Mac Feely, S. Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future? *AI* **2020**, *1*, 143-155.