

Evolving Conversational AI: Analyzing the Development and Advancements in ChatGPT

SUNIL KUMAR SINGH¹, ARTI SACHAN², VARSHA ARYA³

¹ CCET, Panjab University, Chandigarh, India. Email: sksingh@ccet.ac.in

² Insights2Techinfo, USA Email: artisachin019@gmail.com

³ Asia University, Taiwan Email: varshaarya21@gmail.com

ABSTRACT This bibliometric paper presents a comprehensive analysis of a dataset obtained from the Scopus database, spanning the period from 2022 to 2023. The dataset comprises 651 documents sourced from 408 different publications. Through detailed examination, various aspects of the dataset are explored, including document types, authorship patterns, citation metrics, keyword distributions, sources, and country-wise production. The analysis reveals that the dataset consists of recent publications, with an average age of 0.00461 years since publication. Each document, on average, has received 2.584 citations, indicating its impact within the academic community. Collaboration among authors is observed, with an average of 2.61 authors per document and a collaboration index of 3.49, signifying moderate collaboration. The keyword analysis highlights the prominence of terms such as "chatgpt," "artificial intelligence," "chatbot," and "machine learning," indicating a focus on advanced language models and conversational agents. Additionally, the country-wise production analysis identifies significant contributions from the United States, China, the United Kingdom, India, and Australia. This bibliometric study provides valuable insights into the dataset, facilitating a comprehensive understanding of the research landscape. Researchers and stakeholders can utilize these findings to identify emerging trends, forge collaborations, and make informed decisions regarding future research directions. The analysis serves as a valuable resource for exploring the scholarly dynamics within the analyzed field and provides a foundation for further research endeavors.

KEYWORDS Artificial intelligence, Chatbots, Machine learning

I. INTRODUCTION

ChatGPT, a large language model developed by OpenAI, has gained significant attention in recent years due to its remarkable capabilities in generating human-like text responses and engaging in conversational interactions. Powered by artificial intelligence (AI) and machine learning techniques, ChatGPT has found applications in various domains, including customer support, virtual assistants, language translation, and content generation. As the field of AI continues to evolve and language models like ChatGPT become increasingly prevalent, it becomes imperative to gain insights into the research landscape surrounding such technologies [1]–[4].

This study aims to conduct a comprehensive bibliometric analysis of a dataset obtained from the Scopus database, with a particular focus on the prominence of ChatGPT-related research. By examining the dataset, we seek to identify the trends, key themes, and contributions made in the field of ChatGPT and its associated applications. This analysis is crucial to understand the current state of research, explore the interdisciplinary aspects of ChatGPT, and identify emerging areas of interest [5]–[7].

The need for this study arises from the rapid growth and transformative potential of language models like ChatGPT.

With the increasing integration of AI into various industries and sectors, it becomes essential to assess the research landscape surrounding these technologies. By conducting a bibliometric analysis, we can uncover the most relevant keywords, publication sources, and countries involved in ChatGPT-related research. This understanding provides researchers, practitioners, and stakeholders with valuable insights into the advancements, challenges, and future directions of ChatGPT and its applications [8]–[10].

Furthermore, this study aims to contribute to the broader discourse on the ethical considerations and implications associated with large language models. As ChatGPT becomes more sophisticated and capable of generating convincing human-like text, it raises questions regarding the responsible use of AI and the potential impact on areas such as privacy, bias, and misinformation. Through this bibliometric analysis, we can gain insights into the research efforts dedicated to addressing these ethical concerns and fostering a responsible and inclusive development of AI technologies.

In summary, this study seeks to explore the research landscape surrounding ChatGPT and its applications by conducting a comprehensive bibliometric analysis of a dataset obtained from the Scopus database. By analyzing the dataset,

we aim to identify trends, highlight key themes, and contribute to the understanding of the ethical implications and responsible development of large language models like ChatGPT. The findings of this study have the potential to guide future research, foster collaborations, and promote a holistic and informed approach to the utilization of AI technologies in natural language processing and conversation generation.

II. RESEARCH METHODOLOGY

In this study, a systematic research methodology was employed to conduct the bibliometric analysis of the dataset obtained from the Scopus database. The dataset, comprising 651 documents from 408 different sources, was collected within the timespan of 2022 to 2023. To begin, relevant information such as document types, authorship details, citation metrics, and keyword distributions was extracted from the dataset. This extraction process allowed for a comprehensive understanding of the dataset's characteristics.

Subsequently, a rigorous data analysis was conducted to derive insights from the dataset. Descriptive statistics, including averages, frequencies, and percentages, were calculated to examine various aspects such as average years from publication, average citations per document, and average citations per year per document. These statistics provided a quantitative overview of the dataset's citation metrics. Additionally, authorship patterns were analyzed by assessing the average number of authors per document, authors of single-authored versus multi-authored documents, and the collaboration index, which indicates the level of collaboration among authors.

Moreover, the keyword distributions within the dataset were examined to identify prominent and recurring themes. The frequency of keywords such as "chatgpt," "artificial intelligence," "chatbot," and "machine learning" was determined to gauge their significance and relevance within the analyzed documents. This qualitative analysis shed light on the prevalent research topics and areas of interest.

Finally, the dataset's sources and country-wise production were investigated. The frequency of articles from various sources, including prestigious journals and influential publications, was determined to assess their contributions to the dataset. Furthermore, the distribution of research production across countries was analyzed, highlighting the most prolific contributors such as the USA, China, the UK, India, and Australia.

Overall, the research methodology employed in this study encompassed data collection, data extraction, and data analysis. Through this systematic approach, a comprehensive understanding of the dataset was obtained, facilitating the identification of significant patterns, trends, and contributions within the field of study.

III. RESULTS AND ANALYSIS

Fig. 1 obtained from the Scopus database for the purpose of this bibliometric paper encompasses a timespan from 2022 to 2023 and consists of information from 408 different sources,

including journals, books, and other publications. In total, there are 651 documents included in the dataset, with an average age of a mere 0.00461 years since their publication, indicating their recentness. On average, each document has received 2.584 citations and has garnered an average of 2.537 citations per year since its publication, providing insights into its impact over time. The dataset contains a wealth of references, with 11,789 sources cited within the documents.

When it comes to document types, the dataset exhibits a diverse range, including articles, conference papers, conference reviews, editorials, errata, letters, notes, reviews, and short surveys, with varying quantities of each type present. The contents of the documents are enriched with keywords, with 1,574 unique Keywords Plus (ID) and 1,045 unique author's keywords (DE) identified throughout the dataset.

Regarding the authors, a total of 1,700 unique authors are associated with the dataset, collectively making 1,970 appearances. Among them, 190 authors have individually authored documents without any collaboration, while the remaining 1,510 authors have contributed to documents involving multiple authors.

Analyzing the collaboration aspect, the dataset comprises 211 single-authored documents, indicating instances where authors worked independently. On average, each author has contributed to 0.383 documents, and each document has 2.61 authors associated with it. The average number of co-authors per document is 3.03, which, in turn, contributes to a collaboration index of 3.49, indicating a moderate level of collaboration among authors in the dataset.

These comprehensive statistics offer valuable insights into the dataset gathered from the Scopus database, shedding light on the temporal coverage, document types, authorship patterns, citation metrics, and keyword distributions. By leveraging this information, researchers can delve into trends, collaboration dynamics, and research impact within the dataset, facilitating a deeper understanding of the scholarly landscape under investigation.

Fig. 2 comprises articles obtained from various sources, shedding light on the diversity of publications included. Among the sources, the "Annals of Biomedical Engineering" stands out with 35 articles, making it a significant contributor to the dataset. Another prominent source is "Nature," which provided 19 articles. The "Library Hi Tech News" source contributed 13 articles, showcasing its relevance in the dataset. Additionally, several sources such as "Aesthetic Plastic Surgery," the "IEEE/CAA Journal of Automatica Sinica," "Medical Teacher," and "Radiology" each contributed six articles. Other sources, including "Accountability in Research," "Annals of Surgical Oncology," "JMIR Medical Education," and "Journal of Chemical Education," among others, provided varying numbers of articles ranging from five to four. These sources collectively offer a comprehensive range of research topics and perspectives, enriching the dataset with valuable insights from diverse fields of study. Fig 3 provides the information about the country-wise production of articles, providing insights into the distribution of research output.

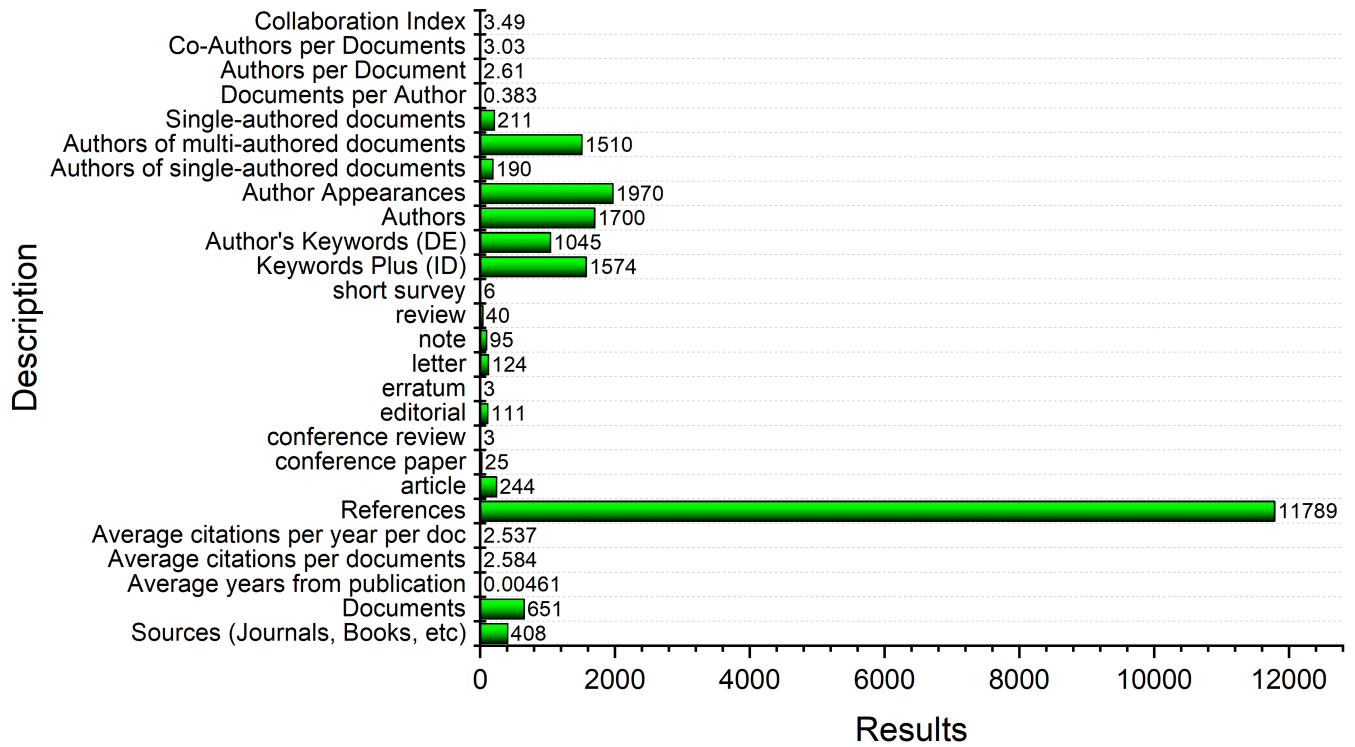


FIGURE 1: Basic Information

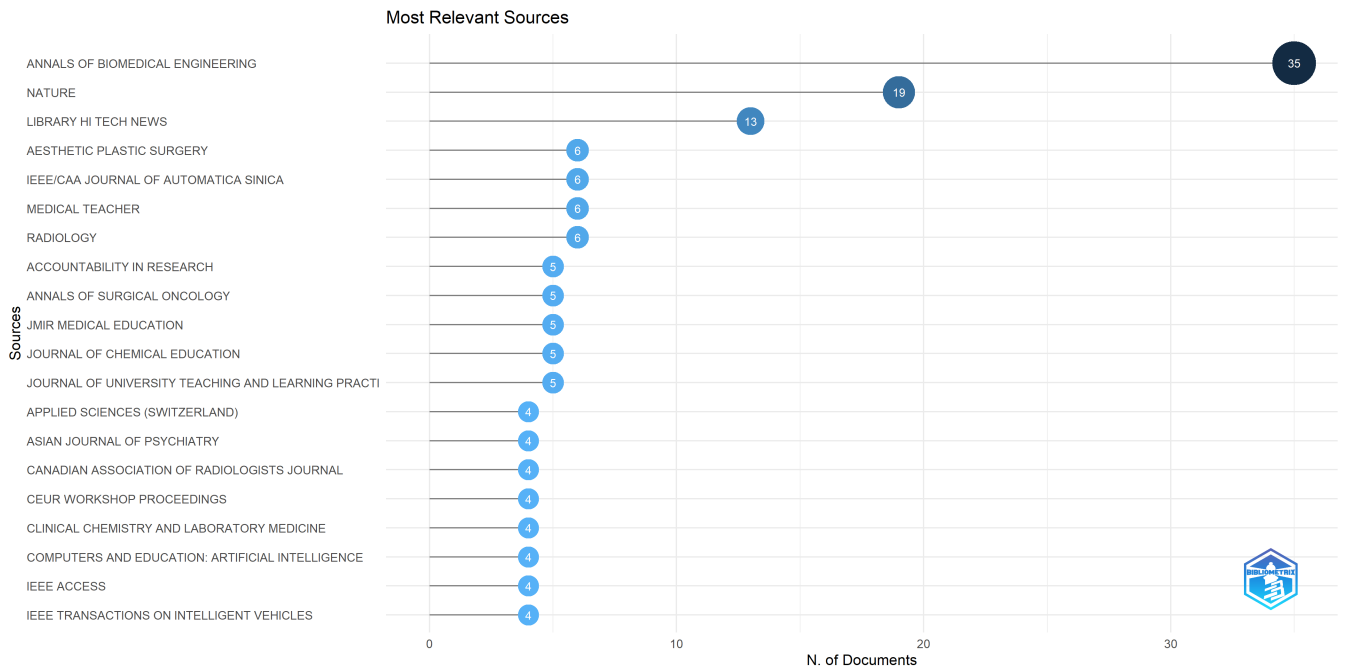


FIGURE 2: Most Relevant Sources

Country Scientific Production

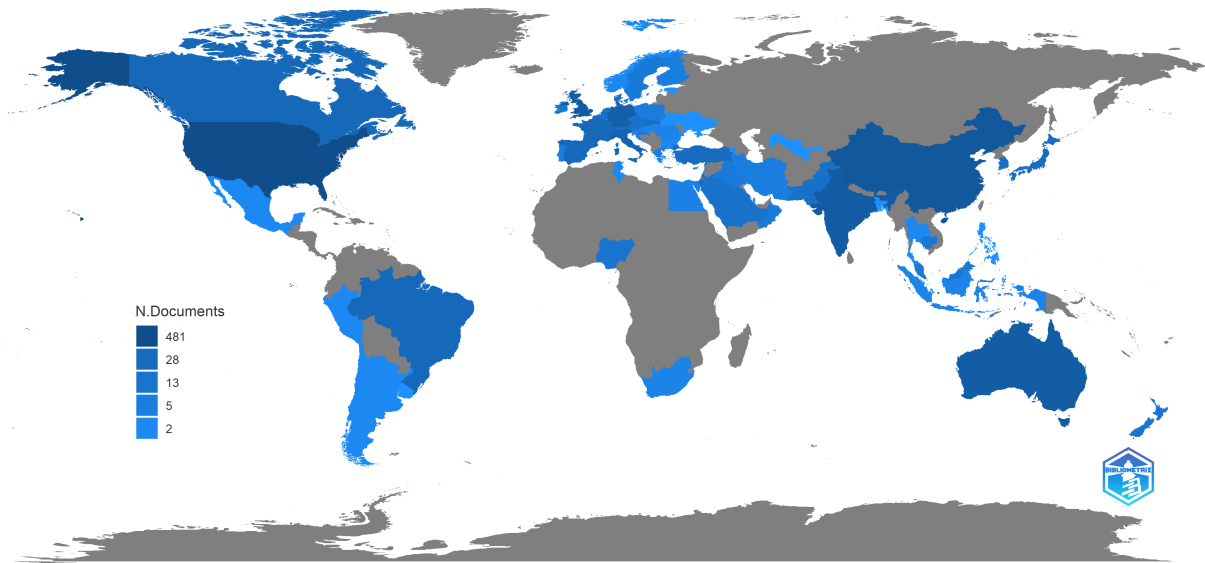


FIGURE 3: Country Scientific Production

The United States (USA) takes the lead with a significant contribution of 481 articles, reflecting its prominence as a major hub of research and scholarly activity. China follows with 179 articles, showcasing its growing research output and impact in the global academic landscape. The United Kingdom (UK) ranks third with 132 articles, highlighting its substantial contributions to the dataset. India and Australia closely follow with 125 and 111 articles, respectively, underscoring their active involvement in research endeavors.

Other notable contributors include Germany with 96 articles, Italy with 69 articles, and Canada and France with 39 articles each. Brazil, Spain, and Singapore also demonstrate considerable research output with 36, 35, and 31 articles, respectively. Switzerland, Israel, Denmark, and Turkey each contributed around 30 articles, further highlighting their involvement in scholarly activities. Japan, Slovakia, South Korea, and the Netherlands follow suit with varying numbers of articles, showcasing their contributions to the dataset.

Overall, the dataset reflects a diverse global representation, with countries from different regions actively engaging in research and knowledge production across various disciplines. These country-wise production statistics provide valuable insights into the distribution and impact of research contributions across different nations, showcasing the global nature of scholarly activities.

The frequency of keywords within the dataset offers a glimpse into the prevalent themes and topics of the articles included (Fig.4). The keyword "chatgpt" appears most frequently, with a frequency count of 254, indicating its significance and relevance in the analyzed documents. "Artificial intelligence" follows closely with a frequency of 172,

highlighting the prominence of this field within the dataset. The keyword "chatbot" appears 40 times, suggesting a focus on the development and utilization of conversational agents. Similarly, "ai" (an abbreviation for artificial intelligence) is mentioned 39 times, emphasizing its pervasive presence in the dataset.

The importance of "machine learning" is also evident, with a frequency of 37, signifying its relevance to the research conducted. "Natural language processing" is another significant keyword, appearing 34 times, which reflects the emphasis on techniques and methodologies for processing and understanding human language within the dataset.

The presence of "large language models" as a keyword, with a frequency of 33, suggests a particular interest in exploring and utilizing advanced language models for various applications. The inclusion of the keyword "ethics" with a frequency of 26 points towards the consideration of ethical implications and concerns surrounding the studied topics.

Other noteworthy keywords include "chatbots" (mentioned 24 times), "education" (mentioned 23 times), and others. The frequencies of these keywords indicate their importance and relevance within the dataset, showcasing the interdisciplinary nature of the research and its potential applications in areas such as education.

Overall, the keyword frequencies provide insights into the key themes, technologies, and areas of focus within the analyzed dataset, offering a glimpse into the prominent keywords that shape the research landscape and drive discussions within the field of study.

pact, and keyword relevance in the academic community. Researchers and stakeholders can utilize these findings to gain a comprehensive overview of the scholarly landscape, identify potential research collaborations, and make informed decisions regarding future research directions. Overall, this study serves as a valuable resource for researchers, academics, and professionals interested in exploring the trends and dynamics of the analyzed dataset from the Scopus database.

REFERENCES

- [1] B. Gupta, S. Gupta, S. Gangwar, M. Kumar, and P. Meena, "Cross-site scripting (xss) abuse and defense: exploitation on several testing bed environments and its defense," *Journal of Information Privacy and Security*, vol. 11, no. 2, pp. 118–136, 2015.
- [2] S. Tripathi and et al., "Hadoop based defense solution to handle distributed denial of service (ddos) attacks," 2013.
- [3] A. Almomani and et al., "Phishing dynamic evolving neural fuzzy framework for online detection zero-day phishing email," arXiv preprint arXiv:1302.0629, 2013.
- [4] M. A. Alsmirat and et al., "Accelerating compute intensive medical imaging segmentation algorithms using hybrid cpu-gpu implementations," *Multimedia Tools and Applications*, vol. 76, pp. 3537–3555, 2017.
- [5] B. B. Gupta, K. Yadav, I. Razzak, K. Psannis, A. Castiglione, and X. Chang, "A novel approach for phishing urls detection using lexical based machine learning in a real-time environment," *Computer Communications*, vol. 175, pp. 47–57, 2021.
- [6] A. P. Plageras and et al., "Efficient large-scale medical data (ehealth big data) analytics in internet of things," in *2017 IEEE 19th Conference on Business Informatics (CBI)*, vol. 2. IEEE, 2017, pp. 21–27.
- [7] K. T. Chui and et al., "An mri scans-based alzheimer's disease detection via convolutional neural network and transfer learning," *Diagnostics*, vol. 12, no. 7, p. 1531, 2022.
- [8] A. Singh and B. B. Gupta, "Distributed denial-of-service (ddos) attacks and defense mechanisms in various web-enabled computing platforms: issues, challenges, and future research directions," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 18, no. 1, pp. 1–43, 2022.
- [9] J. V. Tembhumne and et al., "Mc-dnn: Fake news detection using multi-channel deep neural networks," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 18, no. 1, pp. 1–20, 2022.
- [10] B. Hu, A. Gaurav, C. Choi, and A. Almomani, "Evaluation and comparative analysis of semantic web-based strategies for enhancing educational system development," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 18, no. 1, pp. 1–14, 2022.
- [11] H. H. Thorp, "Chatgpt is fun, but not an author," *Science*, vol. 379, no. 6630, p. 313, 2023, all Open Access, Bronze Open Access.
- [12] C. Stokel-Walker, "Chatgpt listed as author on research papers: many scientists disapprove," *Nature*, vol. 613, no. 7945, p. 620 – 621, 2023.
- [13] E. A. M. van Dis, J. Bollen, W. Zuidema, R. van Rooij, and C. L. Bockting, "Chatgpt: five priorities for research," *Nature*, vol. 614, no. 7947, p. 224 – 226, 2023.
- [14] "Tools such as chatgpt threaten transparent science; here are our ground rules for their use," *Nature*, vol. 613, no. 7945, p. 612, 2023, all Open Access, Bronze Open Access.
- [15] H. Else, "Abstracts written by chatgpt fool scientists," *Nature*, vol. 613, no. 7944, p. 423, 2023, all Open Access, Bronze Open Access.
- [16] Y. Shen and et al., "Chatgpt and other large language models are double-edged swords," *Radiology*, vol. 307, no. 2, 2023.
- [17] C. Stokel-Walker, "Ai bot chatgpt writes smart essays — should academics worry?" *Nature*, 2022.
- [18] C. Stokel-Walker and R. Van Noorden, "What chatgpt and generative ai mean for science," *Nature*, vol. 614, no. 7947, p. 214 – 216, 2023.
- [19] A. Gilson, C. W. Safranek, T. Huang, V. Socrates, L. Chi, R. A. Taylor, and D. Chartash, "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*, vol. 9, 2023, all Open Access, Gold Open Access, Green Open Access.
- [20] S. Biswas, "Chatgpt and the future of medical writing," *Radiology*, vol. 307, no. 2, 2023.
- [21] M. Liebrez, R. Schleifer, A. Buadze, D. Bhugra, and A. Smith, "Generating scholarly content with chatgpt: ethical challenges for medical publishing," *The Lancet Digital Health*, vol. 5, no. 3, p. e105 – e106, 2023, all Open Access, Gold Open Access, Green Open Access.
- [22] S. B. Patel and K. Lam, "Chatgpt: the future of discharge summaries?" *The Lancet Digital Health*, vol. 5, no. 3, p. e107 – e108, 2023, all Open Access, Gold Open Access.
- [23] J. V. Pavlik, "Collaborating with chatgpt: Considering the implications of generative artificial intelligence for journalism and media education," *Journalism and Mass Communication Educator*, vol. 78, no. 1, p. 84 – 93, 2023.
- [24] Y. K. Dwivedi and et al., "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*, vol. 71, 2023.
- [25] M. Salvagno, F. S. Taccone, and A. G. Gerli, "Can artificial intelligence help for scientific writing?" *Critical Care*, vol. 27, no. 1, 2023, all Open Access, Gold Open Access, Green Open Access.
- [26] B. Gordijn and H. t. Have, "Chatgpt: evolution or revolution?" *Medicine, Health Care and Philosophy*, vol. 26, no. 1, p. 1 – 2, 2023, all Open Access, Bronze Open Access.
- [27] F.-Y. Wang, Q. Miao, X. Li, X. Wang, and Y. Lin, "What does chatgpt say: The dao from algorithmic intelligence to linguistic intelligence," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 3, p. 575 – 579, 2023, all Open Access, Bronze Open Access.
- [28] E. Kasneci and et al., "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, 2023, all Open Access, Bronze Open Access, Green Open Access.
- [29] M. Sallam, "Chatgpt utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns," *Healthcare (Switzerland)*, vol. 11, no. 6, 2023, all Open Access, Gold Open Access, Green Open Access.
- [30] F. C. Kitamura, "Chatgpt is shaping the future of medical writing but still requires human judgment," *Radiology*, vol. 307, no. 2, 2023.