ORIGINAL

# Unveiling the Thematic Landscape of Generative Pre-trained Transformer (GPT) Through Bibliometric Analysis

## Redes de conocimiento y colaboración internacional en torno al Generative Pre-trained Transformer (GPT): Un estudio bibliométrico

Carlos Alberto Gómez Cano[1] ✉, Verenice Sánchez Castillo[2], Tulio Andrés Clavijo Gallego[3]

[1]Corporación Unificada Nacional de Educación Superior - CUN, Florencia, Colombia.
[2]Universidad de la Amazonía, Florencia, Colombia.
[3]Universidad del Cauca, Popayán, Colombia.

**ABSTRACT**

**Introduction:** the Generative Pre-trained Transformer (GPT) is a deep learning language model architecture developed by OpenAI.
**Aim:** to describe the knowledge networks (both at the theoretical and country levels) of the Generative Pre-trained Transformer (GPT) as an emerging technology.
**Results:** 222 Documents were identified, of which 69 were articles, 50 were conference papers, 36 were editorials, 29 were notes, 19 were letters, 14 were reviews, 3 were conference reviews, and 2 were short surveys. In terms of the number of documents per year, 2 were found in 2019, 10 in 2020, 22 in 2021, 44 in 2022, and 144 in 2023. The year-on-year growth rate was over 100 % in all years. The subject area with the highest number of documents was Computer Science with 90 documents. The most productive countries in relation to GPT were the United States with 60 documents, followed by China with 19, the United Kingdom with 18, India with 15, and Australia with 12. Co-occurrence illustrated the centrality of Artificial Intelligence, Natural Language Processing, Deep Learning, and the term Human around ChatGPT and GPT.
**Conclusions:** this bibliometric study aimed to describe the knowledge networks of the Generative Pre-trained Transformer (GPT) as an emerging technology. Although only 222 documents were found, this study revealed a high level of international scientific collaboration in the field. The results suggest that GPT is a highly relevant technology with a wide range of potential applications in natural language processing, artificial intelligence, and deep learning.
Moreover, the study was able to qualitatively characterize the main thematic areas surrounding GPT, including its applications in chatbots, text generation, machine translation, sentiment analysis, and more.

**Keywords:** Generative Pre-trained Transformer; GPT; GPT Chat; Natural Language Processing; Artificial Intelligence.

**RESUMEN**

**Introducción:** el Generative Pre-trained Transformer (GPT) es una arquitectura de modelo de lenguaje de aprendizaje profundo desarrollada por OpenAI.
**Objetivo:** describir las redes de conocimiento (tanto a nivel teórico como de país) del Generative Pre-trained Transformer (GPT) como una tecnología emergente.
**Resultados:** se identificaron 222 documentos, de los cuales 69 eran artículos, 50 eran ponencias de conferencias, 36 eran editoriales, 29 eran notas, 19 eran cartas, 14 eran reseñas, 3 eran reseñas de conferencias y 2 eran encuestas cortas. En términos del número de documentos por año, se encontraron 2 en 2019, 10 en 2020, 22

los años. El área temática con el mayor número de documentos fue Ciencias de la Computación con 90 documentos. Los países más productivos en relación al GPT fueron los Estados Unidos con 60 documentos, seguidos de China con 19, el Reino Unido con 18, India con 15 y Australia con 12. La co-ocurrencia ilustró la centralidad de la Inteligencia Artificial, el Procesamiento de Lenguaje Natural, Deep Learning y el término Human alrededor de ChatGPT y GPT.

**Conclusiones:** este estudio bibliométrico tuvo como objetivo describir las redes de conocimiento del Generative Pre-trained Transformer (GPT) como una tecnología emergente. Aunque solo se encontraron 222 documentos, este estudio reveló un alto nivel de colaboración científica internacional en el campo. Los resultados sugieren que GPT es una tecnología altamente relevante con una amplia gama de aplicaciones potenciales en procesamiento de lenguaje natural, inteligencia artificial y aprendizaje profundo. Además, el estudio pudo caracterizar cualitativamente las principales áreas temáticas que rodean al GPT, incluyendo sus aplicaciones en chatbots, generación de texto, traducción automática, análisis de sentimientos y más. en 2021, 44 en 2022 y 144 en 2023. La tasa de crecimiento interanual fue superior al 100 % en todos

**Palabras clave:** Generative Pre-trained Transformer; GPT; GPT Chat; Procesamiento de Lenguaje Natural; Inteligencia Artificial.

## INTRODUCTION

The Generative Pre-trained Transformer (GPT) is a deep learning language model architecture developed by OpenAI. The latest version of GPT is GPT-4, which was released in 2023 and is one of the largest and most powerful language models that exist, with 100 trillion parameters, nearly 600 times more than its predecessor. [1]

The current relevance of GPT-4 and the GPT architecture in general is very high due to its ability to generate natural and coherent text in a wide variety of natural language processing tasks. This tool has been used in a large number of applications, including chatbots, text generation, automatic translation, sentiment analysis, and more. [1]

In addition, GPT has driven research in the field of deep learning and has led to new innovations and improvements in language model architecture. GPT's ability to generate high-quality text and its adaptability to a wide variety of natural language processing tasks has led to the creation of new applications and opportunities in the field of artificial intelligence. [2,3]

The objective of this bibliometric study is primarily focused on describing the knowledge networks (both at the theoretical and country levels) of the Generative Pre-trained Transformer (GPT) as an emerging technology.

## METHODS

The bibliometric study utilized Scopus as the database for collecting relevant publications related to the Generative Pre-trained Transformer (GPT) as an emerging technology. Scopus is one of the largest abstract and citation databases of peer-reviewed literature, providing a comprehensive overview of research output in various fields.

*Data Collection*

A search was conducted on Scopus using the following search string: TITLE-ABS-KEY("Generative Pre-trained Transformer" OR "ChatGPT" OR "GptChat" OR "Chat GPT" OR "GPT Chat").

*Data Analysis*

The retrieved data was analyzed using the bibliometric software Vosviewer. Vosviewer is a popular bibliometric analysis tool that allows for the visualization and analysis of bibliometric data, including the generation of maps of science and network diagrams.

The bibliometric analysis focused on the identification of key authors and countries that have contributed to the literature on GPT.

Network analysis was conducted to visualize the co-occurrence of keywords in the GPT literature. The analysis was used to identify clusters of related keywords and to explore the relationships between different areas of research related to GPT.

## RESULTS

222 documents were identified, of which 69 were articles, 50 were conference papers, 36 were editorials, 29 were notes, 19 were letters, 14 were reviews, 3 were conference reviews, and 2 were short surveys. In terms of the number of documents per year, 2 were found in 2019, 10 in 2020, 22 in 2021, 44 in 2022, and 144 in 2023.

The year-on-year growth rate was over 100 % in all years (figure 1).
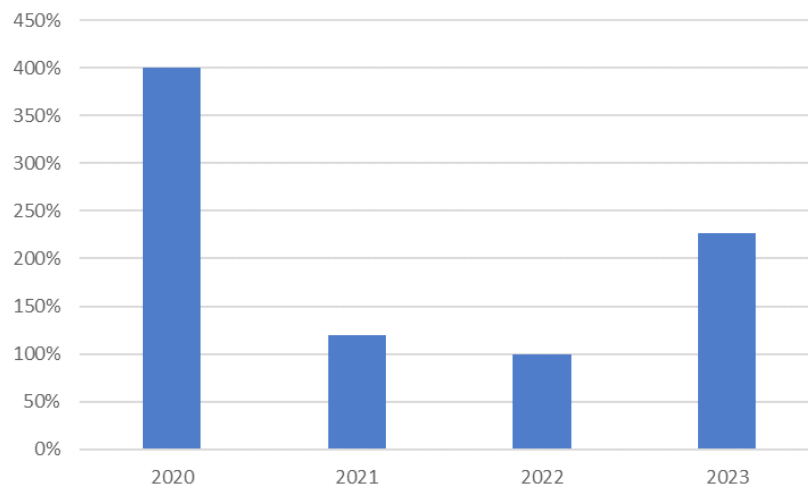


**Figure 1.** Growth rate

The subject area with the highest number of documents was Computer Science with 90 documents, followed by Medicine with 61. table 1 shows the distribution of documents by subject area.

| Table 1. Documents by Subject area | |
|---|---|
| **Subject area** | **Ndoc** |
| Computer Science | 90 |
| Medicine | 61 |
| Social Sciences | 56 |
| Engineering | 40 |
| Mathematics | 24 |
| Decision Sciences | 20 |
| Multidisciplinary | 18 |
| Arts and Humanities | 13 |
| Health Professions | 13 |
| Business, Management and Accounting | 10 |
| Biochemistry, Genetics and Molecular Biology | 8 |
| Nursing | 8 |
| Physics and Astronomy | 8 |
| Immunology and Microbiology | 7 |
| Psychology | 7 |
| Chemical Engineering | 6 |
| Environmental Science | 5 |
| Materials Science | 5 |
| Neuroscience | 5 |
| Chemistry | 3 |
| Energy | 3 |
| Economics, Econometrics and Finance | 2 |
| Agricultural and Biological Sciences | 1 |
| Earth and Planetary Sciences | 1 |

The most productive countries in relation to GPT were the United States with 60 documents, followed by China with 19, the United Kingdom with 18, India with 15, and Australia with 12. figure 2 displays the

collaboration network among countries, where five country cores can be observed, with the United States having the highest centrality. On the other hand, figure 3 shows the co-authorship networks.
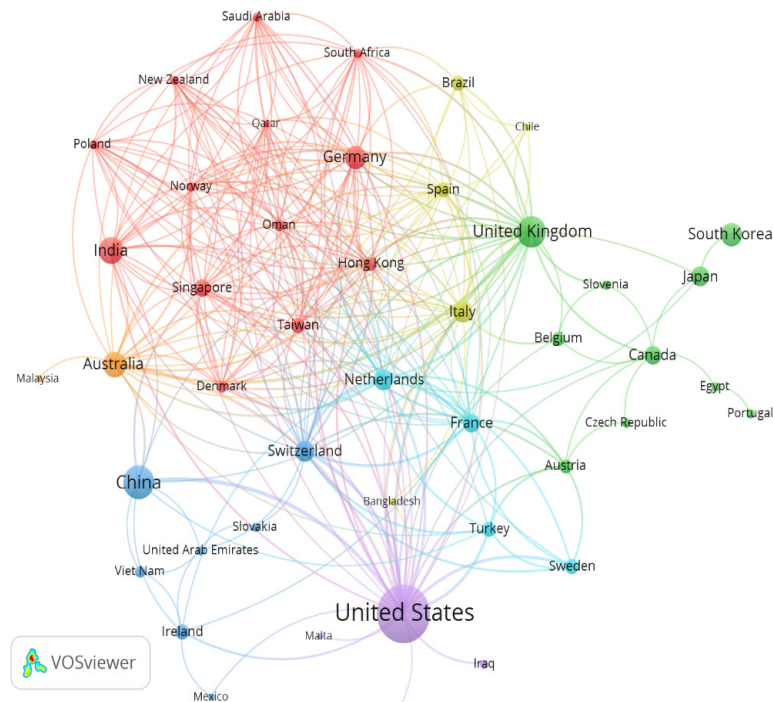


**Figure 2.** Collaboration networks between countries
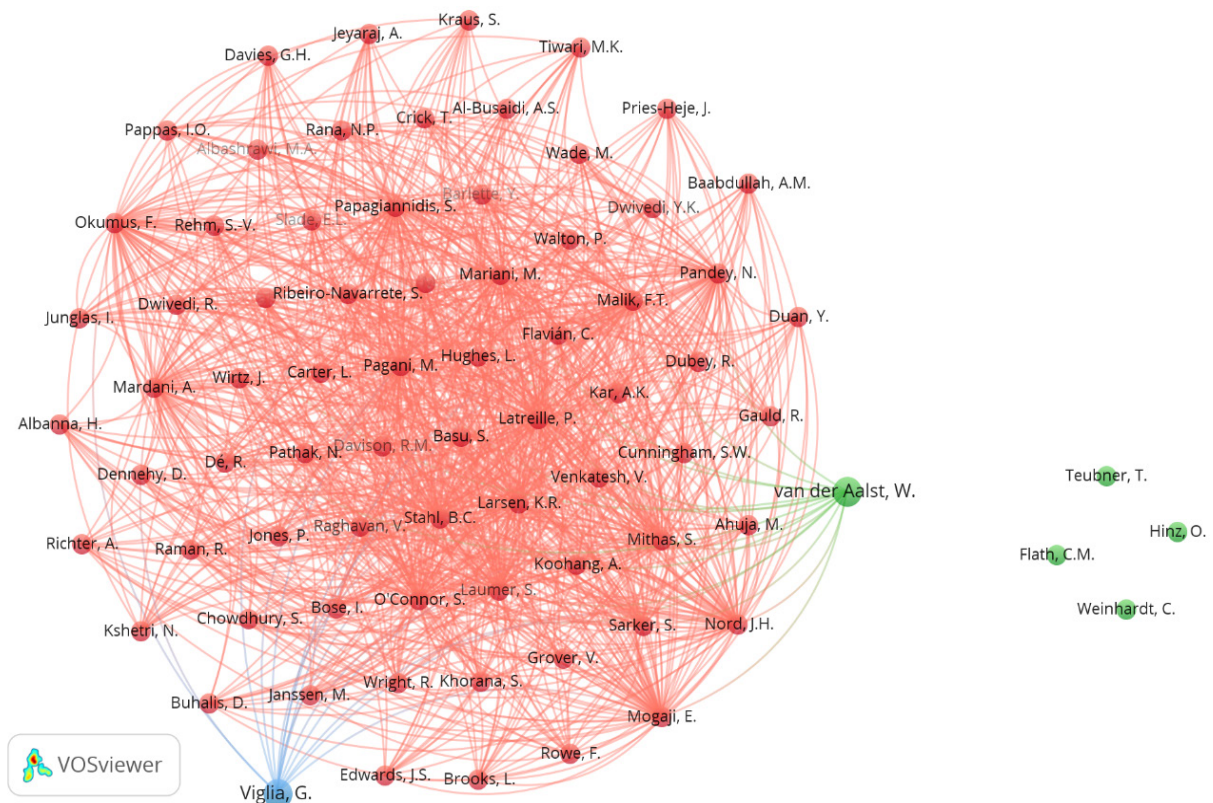


**Figure 3.** Co-authorship networks

Both figure 4 (density map) and figure 5 (term co-occurrence) illustrate the centrality of Artificial Intelligence, Natural Language Processing, Deep Learning, and the term Human around ChatGPT and GPT.
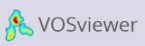
**Figure 4.** Terms Density map



**Figure 5.** Term co-occurrence

## DISCUSSION

GPT is a powerful tool in the field of natural language processing, and its impact on scientific and technological development cannot be overstated. Its ability to generate natural and coherent text in response to user input has been a game-changer in the development of chatbots and virtual assistants.

These applications have become increasingly popular in recent years, and the development of ChatGPT has played a crucial role in their success. By allowing for more natural and engaging conversations between humans and machines, ChatGPT has opened up a world of new possibilities for the use of virtual assistants in a wide

range of industries.

Moreover, the development of ChatGPT has also had a significant impact on the field of artificial intelligence as a whole. Its use of deep learning techniques and advanced algorithms has led to new breakthroughs in the development of language models and other AI applications. As researchers continue to refine and improve ChatGPT, it is likely that we will see even more impressive developments in the field of AI in the years to come. This technology has already proven to be a powerful tool for advancing scientific and technological progress, and its potential for future innovations is virtually limitless. [4,5,6]

In addition to its practical applications, ChatGPT has also had an important impact on the way we think about language and communication. By simulating human-like conversations, ChatGPT has helped to shed new light on the complexities of human language and the ways in which we use it to convey meaning. As researchers continue to study and refine ChatGPT and other natural language processing tools, we are likely to gain a deeper understanding of the way language works and the ways in which it shapes our interactions with the world around us. This knowledge has the potential to inform a wide range of fields, from linguistics and psychology to philosophy and beyond. [7,8,9]

Several countries have been actively involved in the development of Generative Pre-trained Transformer (GPT) technology. The United States is the most productive country, with 60 documents related to GPT found in the Scopus database. The country has been at the forefront of artificial intelligence research, with major companies like OpenAI, Google, and Microsoft headquartered there. China is the second most productive country, with 19 GPT-related documents found in Scopus. The country has been investing heavily in AI research and development, and its companies are known for their advancements in natural language processing and machine learning.

The United Kingdom and India are also prominent contributors to GPT-related research, with 18 and 15 documents found in Scopus, respectively. The UK is known for its strong academic tradition and has several top-ranked universities that have been involved in GPT research. India, on the other hand, has a large tech industry and is home to several companies that have been developing AI and machine learning technologies. Australia is another country that has been active in GPT research, with 12 documents found in Scopus. The country has a strong research focus on natural language processing and has several universities that have been involved in GPT-related research.

The potential uses of natural language processing based on artificial intelligence in scientific research are vast and promising. One of the most significant applications is the ability to analyze and extract valuable insights from large amounts of unstructured data. With AI-powered tools such as chatbots and virtual assistants, researchers can efficiently interact with data sources, extract relevant information, and gain deeper insights into complex scientific problems. Furthermore, AI can help automate routine tasks such as data cleaning and data analysis, freeing up researchers' time to focus on more critical tasks. [10,11,12,13]

The studies found in the bibliometric analysis indicate that the main themes of research in the field of natural language processing based on AI include artificial intelligence, natural language processing, deep learning, and human. These themes suggest that researchers are interested in exploring how AI can be used to improve natural language processing technologies and how these technologies can be used to enhance human interaction with technology. For example, AI-powered chatbots and virtual assistants can provide personalized recommendations to users, assist in language learning, and provide customer service. Moreover, AI can help researchers better understand how humans interact with language, opening up new avenues of research in fields such as psychology and linguistics.

The potential uses of natural language processing based on AI in scientific research are vast and promising. [14] The studies found in the bibliometric analysis indicate that research in this field is focused on improving AI-powered technologies and understanding how these technologies can be used to enhance human interaction with technology. As AI continues to evolve and become more advanced, it is likely that we will see even more innovative uses of natural language processing in scientific research.

*Limitations*

One potential limitation of the study is the use of Scopus as the only database. Scopus covers a large number of journals, but it is possible that some relevant publications may have been missed. Additionally, the study only covers publications up to early 2023, and new research may have been published since then.

## CONCLUSIONS

In conclusion, this bibliometric study aimed to describe the knowledge networks of the Generative Pre-trained Transformer (GPT) as an emerging technology. Although only 222 documents were found, this study revealed a high level of international scientific collaboration in the field. The results suggest that GPT is a highly relevant technology with a wide range of potential applications in natural language processing, artificial intelligence, and deep learning.

Moreover, the study was able to qualitatively characterize the main thematic areas surrounding GPT, including its applications in chatbots, text generation, machine translation, sentiment analysis, and more. The findings suggest that GPT has significant potential to drive innovation and contribute to the advancement of the field of artificial intelligence.

Overall, despite the limited number of documents found, this study provides valuable insights into the knowledge networks and potential applications of GPT. The high level of international collaboration and the identified thematic areas suggest that GPT is a promising area for future research and development.

## REFERENCES

1.    OpenAI. GPT-4. OpenAI 2023. https://openai.com/product/gpt-4.

2.    Castillo-Gonzalez W. ChatGPT and the future of scientific communication. Metaverse Basic Appl Res 2022;1:8. https://doi.org/10.56294/mr20228.

3.    Castillo-González W. The importance of human supervision in the use of ChatGPT as a support tool in scientific writing. Metaverse Basic Appl Res 2023;2:29. https://doi.org/10.56294/mr202329.

4.    Lund BD, Wang T. Chatting about ChatGPT: how may AI and GPT impact academia and libraries? Libr Hi Tech News 2023;ahead-of-print. https://doi.org/10.1108/LHTN-01-2023-0009.

5.    McGee RW. Is Chat Gpt Biased Against Conservatives? An Empirical Study 2023. https://doi.org/10.2139/ssrn.4359405.

6.    Biswas SS. Role of Chat GPT in Public Health. Ann Biomed Eng 2023. https://doi.org/10.1007/s10439-023-03172-7.

7.    Cowen T, Tabarrok AT. How to Learn and Teach Economics with Large Language Models, Including GPT 2023. https://doi.org/10.2139/ssrn.4391863.

8.    Biswas S. ChatGPT and the Future of Medical Writing. Radiology 2023:223312. https://doi.org/10.1148/radiol.223312.

9.    Biswas SS. Potential Use of Chat GPT in Global Warming. Ann Biomed Eng 2023. https://doi.org/10.1007/s10439-023-03171-8.

10.    Baclic O, Tunis M, Young K, Doan C, Swerdfeger H, Schonfeld J. Challenges and opportunities for public health made possible by advances in natural language processing. Can Commun Dis Rep 2020;46:161–8. https://doi.org/10.14745/ccdr.v46i06a02.

11.    Reshamwala A, Mishra D, Pawar P. REVIEW ON NATURAL LANGUAGE PROCESSING. IRACST – Eng Sci Technol Int J ESTIJ 2013;3:113–6.

12.    Wu L, Dodoo NA, Wen TJ, Ke L. Understanding Twitter conversations about artificial intelligence in advertising based on natural language processing. Int J Advert 2022;41:685–702. https://doi.org/10.1080/02650487.2021.1920218.

13.    Juhn Y, Liu H. Artificial intelligence approaches using natural language processing to advance EHR-based clinical research. J Allergy Clin Immunol 2020;145:463–9. https://doi.org/10.1016/j.jaci.2019.12.897.

14.    Cano CAG, Castillo VS, Gallego TAC. Mapping the Landscape of Netnographic Research: A Bibliometric Study of Social Interactions and Digital Culture. Data Metadata 2023;2:25. https://doi.org/10.56294/dm202325.

## CONFLICT OF INTEREST

No conflict of interest.

**AUTHORSHIP CONTRIBUTION**

*Conceptualization:* Carlos Alberto Gómez Cano, Verenice Sánchez Castillo, Tulio Andrés Clavijo Gallego.

*Methodology:* Carlos Alberto Gómez Cano, Verenice Sánchez Castillo, Tulio Andrés Clavijo Gallego.

*Software:* Carlos Alberto Gómez Cano, Verenice Sánchez Castillo, Tulio Andrés Clavijo Gallego.

*Investigation:* Carlos Alberto Gómez Cano, Verenice Sánchez Castillo, Tulio Andrés Clavijo Gallego.

*Original writing-drafting:* Carlos Alberto Gómez Cano, Verenice Sánchez Castillo, Tulio Andrés Clavijo Gallego.

*Writing-revision and editing:* Carlos Alberto Gómez Cano, Verenice Sánchez Castillo, Tulio Andrés Clavijo Gallego.