

# ChatGPT Research: Insights from Early Studies Using Network Scientometric Approach

Hiran H. Lathabai<sup>1,\*</sup>, Thara Prabhakaran<sup>2</sup>, Raghu Raman<sup>3</sup>

<sup>1</sup>Amrita CREATE, Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, INDIA.

<sup>2</sup>Department of Futures Studies, University of Kerala, Thiruvananthapuram, Kerala, INDIA.

<sup>3</sup>Amrita School of Business, Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, INDIA.

## ABSTRACT

The introduction of generative AI models, especially OpenAI's ChatGPT, has profoundly impacted several fields. To uncover key fields of research, key research clusters, emerging research topics, key research contributions within grown and emerging clusters, and key insightful implications for various stakeholders, this study analyses the early body of scientific literature ( $n=1873$ ) related to ChatGPT research indexed in Dimensions (from November 29, 2022 to May 20, 2022) database using network scientometric approach. This approach employs network mining of two major networks related to scientific literature for knowledge discovery exercise. Science mapping analysis using the Fields of Research (FoRs) network revealed key fields of research impacted by ChatGPT. Scientific literature mining is conducted using publications citation network analysis with the help of the Flow Vergence model and cluster analysis. Major growth clusters that contributed and might continue significantly to the network's growth are identified and found to be associated with education in general, medical education, medical diagnosis and clinical writing, scientific writing, and systematic literature review. Important emerging clusters are found to mostly deal with novel applications like harmful content detection in social media, annotation, assessment, etc., Further, the dynamics of grown clusters and emerging clusters was tracked on Dec 31, 2023 (after six months). All the clusters are found to have grown significantly. Cluster merging is witnessed in the case of grown clusters, making the new clusters multi-themed and overwhelmed by incremental contributions. However, the merger of emerging clusters contributed to the formation of relatively better-performing clusters. Through this knowledge discovery exercise, the paper highlights the knowledge and technology advancement of ChatGPT and its potential in numerous fields and sheds light on pressing problems and the moral dilemmas raised by its use. The analysis reveals several policy implications for various stakeholders, including education and research policymakers.

**Keywords:** ChatGPT, Generative AI, Flow Vergence, FV index, Science mapping, Scientific literature mining.

## Correspondence:

**Hiran H. Lathabai**

Amrita CREATE, Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, INDIA.  
Email: hiranhl007@gmail.com  
ORCID: 0000-0002-5633-9842

**Received:** 14-02-2024;

**Revised:** 04-07-2024;

**Accepted:** 16-08-2024.

## INTRODUCTION

Artificial Intelligence (AI) is evolving quickly and has impacted many research and industrial practice fields. A recent breakthrough in the form of sophisticated language models like ChatGPT, which can produce text responses similar to those of humans,<sup>[1]</sup> may accelerate the invasion of AI into several unprecedented areas. Let us briefly examine the evolution of AI language models. The roots of AI language models can be traced back to the 1950s and 1960s when researchers like Alan Turing and Noam Chomsky laid the foundation for computational linguistics and the theory of formal languages.<sup>[2]</sup> However, it was

not until the 1980s that the first practical language processing systems, such as ELIZA and SHRDLU, were developed.<sup>[3]</sup> These early systems utilized rule-based approaches, where language understanding and generation were based on predefined rules and limited in their capabilities and scalability. AI language models have seen significant milestones, leading to more sophisticated and robust models.<sup>[4]</sup> In the 1990s, statistical language models based on probabilistic methods, such as n-gram models, became popular. Compared to rule-based approaches, these models could learn from large text data corpora and generate more fluent and coherent text.<sup>[5]</sup> In the late 2000s and early 2010s, neural network-based language models started gaining attention. Models like the Recursive Neural Network (RNN) and the Long Short-Term Memory (LSTM) networks demonstrated improved language modeling capabilities compared to statistical models. These paved the way for more advanced models.<sup>[6,7]</sup>



DOI: 10.5530/jscires.20041177

### Copyright Information :

Copyright Author (s) 2024 Distributed under  
Creative Commons CC-BY 4.0

Publishing Partner : Manuscript Technomedia. [www.mstechnomedia.com]

The introduction of the Transformer model by Vaswani *et al.*,<sup>[8]</sup> revolutionized the field of AI language models. The Transformer architecture, based on self-attention mechanisms, allows for more efficient and parallelizable training of large-scale language models, leading to the development of models like GPT-2 and GPT-3, which achieved state-of-the-art performance in various language-related tasks.<sup>[9]</sup> The availability of large-scale text data, advancements in hardware, and deep learning techniques have been critical enablers in the development of ChatGPT.<sup>[10]</sup> The GPT-3.5 architecture-based ChatGPT big language model was created by OpenAI and trained on a sizable corpus of online content.<sup>[11]</sup>

Since its debut, ChatGPT has attracted much interest and has been used in various contexts, such as chatbots, virtual assistants, content creation, and more. It is critical to comprehend the scientific research surrounding each emergent technology. As Cooper<sup>[12]</sup> pointed out, research related to ChatGPT needs thorough evaluation. Though attempts to review ChatGPT for its capability, shortfalls, etc., in various fields of applications can be found, a systematic analysis of the body of literature related to ChatGPT, especially concerning aspects like 1) key contributing fields of research, 2) key contributing clusters formed in ChatGPT literature and their growth prospects, 3) emerging clusters that need attention, 4) major specific contributions within grown and emerging clusters, etc., is not found according to the best of our knowledge. Such an analysis can help determine useful insights for various stakeholders. This gap is addressed in this work.

The field of *Scientometrics* deals with the study of quantitative methods of research on the development of science as an informational process.<sup>[13]</sup> A scientometric analysis is useful for learning about research trends, patterns, and effects in a study area. The network approach to scientometrics or network scientometric approach,<sup>[14]</sup> uses various networks associated with scientific literature for mining scientific literature for knowledge discovery exercises, especially for extracting key insightful implications for various stakeholders, including policymakers. It is critical to examine the scholarly output associated with ChatGPT to pinpoint the trends and patterns formed in the area because ChatGPT has attracted considerable attention and has been widely adopted by researchers and practitioners alike. In the case of ChatGPT, an early scientometric analysis of the scientific literature related to ChatGPT can offer insightful information about the early developments and trends in ChatGPT research that might reveal the overall influence of ChatGPT on the scientific community. This prompted the current investigation which is largely based on early scientometric analysis, the objectives of which can be stated as:

1. Determination of key Fields of Research (FoRs) formed by the science mapping behind the ChatGPT literature at its early stage.
2. Determination of key research clusters formed within ChatGPT literature in its early stage and state of knowledge flow vergence exhibited by these early stage clusters.
3. Determination of emerging research topics related to ChatGPT by identifying emerging research clusters at its early stage.
4. Determination of key research contributions within substantially grown and emerging clusters at the early stage.
5. Tracking the dynamics of grown clusters and emerging clusters so as to,
6. Extrication of key insights that will be helpful for several stakeholders, including policymakers. The potential benefits of this investigation are briefly discussed next. A systematically conducted scientometric analysis pursuing the above-stated objectives can offer a quantitative and qualitative understanding of the research output related to ChatGPT. Additionally, by identifying potential research gaps or areas that need more study, this analysis can help guide future research efforts and make it easier to create new ChatGPT applications and use cases. Also, a thorough awareness of the research related to ChatGPT can aid in the decision-making process for academics, practitioners, and policymakers regarding the course and potential outcomes of others as well as their own research or development initiatives. This study can inform the above-listed as well as stakeholders. Details of the methodology of the investigation are discussed next.

## METHODOLOGY

### Network scientometric approach for early scientometric analysis

The body of knowledge represented by scientific literature is associated with the relationship between so many actors. These relationships can be represented as networks. A network is a construct with an underlying graph (set of vertices and links between these vertices) and information about both of these sets. Due to the existence of a large number of these kinds of relationships in scientific literature and the possibility of a network representation of these, a vast plethora of opportunities are there for network-based analysis of scientific literature, especially for scientometric applications. The umbrella of network-based methods, tools, and techniques useful for scientometric applications can be termed as the 'network approach to scientometrics' or 'network scientometric' approach.<sup>[14]</sup> Due to this potential of the network-scientometric approach, the same is adopted for addressing the above-mentioned objectives. A concise note on how the objectives can be achieved by using

relevant network representations relationships within the body of scientific literature related to a chosen topic (in this case, 'ChatGPT') is discussed first before discussing the details of the methodology.

As the first objective of this study is to determine the key Fields of Research (FoRs) related to ChatGPT research, mapping of the relationship between different FoRs associated with ChatGPT literature has to be done. However, the relationship among different FoRs is not a direct relationship. FoRs are indirectly related to each other through the number of common publications directly associated with each FoR. Thus, to determine the (indirect) mapping among FoRs, the (direct) mapping of publications within the ChatGPT literature to different FoRs is required. This requires the creation of an affiliation or 2-mode network that represents the relationship between publications and FoRs. From this affiliation network, a co-affiliation network of FoRs that reflects the indirect relationship among FoRs can be constructed. This network can be analyzed with useful network analysis techniques (which will be discussed shortly) to determine the FoRs that occupy key positions in the FoR map/network. This science mapping exercise on ChatGPT literature can address our first objective.

The next three objectives are related to different aspects of the knowledge flow relationship between publications in the ChatGPT literature. Hence, a mapping that reflects the knowledge flow between the publications is required and hence the creation of citation networks (with publications as vertices

and citation links as directed edges) is the best way to carry out such investigations. As the determination of key research clusters within ChatGPT literature is the second objective, the ChatGPT literature (represented by citation network) needs to be partitioned into different clusters using suitable cluster formation algorithms (which will be shortly discussed). This can give an idea about substantially grown clusters (partially addresses the second objective) and not-so-grown (emerging) clusters (third objective). The knowledge flow among publications is the key aspect that determines the present growth as well as the growth potential of research clusters and also the literature (related to a topic) as a whole. Hence, an existing model that can reflect the knowledge flow characteristics within citation networks viz. the Flow Vergence or FV model<sup>[15]</sup> can be used together with cluster analysis, for determining the key grown research clusters (to address the second objective completely). Once key grown clusters and emerging clusters are determined, important major specific contributions within these clusters can be determined. For that key publications within each of these clusters have to be determined (the FV model can help in this regard too). Then, the contributions of these publications can be determined through content analysis (the fourth objective). For tracking the progress of grown and emerging clusters (objective 5), data is collected after an interval of suitable duration using the same search term from the same database. Steps required to achieve objectives 3 and 4 are repeated with the following modification. As finding out the performance of all the existing clusters after the specific time interval is beyond the scope (as we are tracking the growth

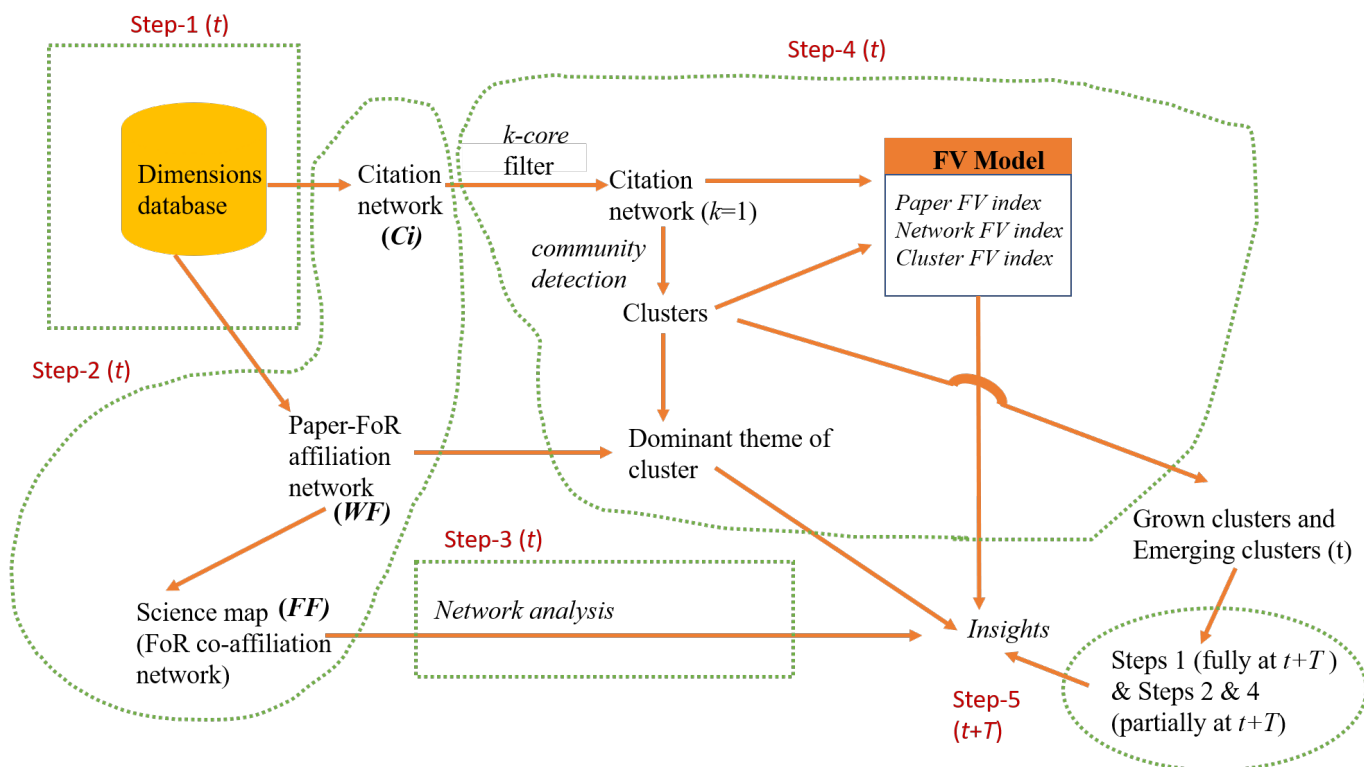
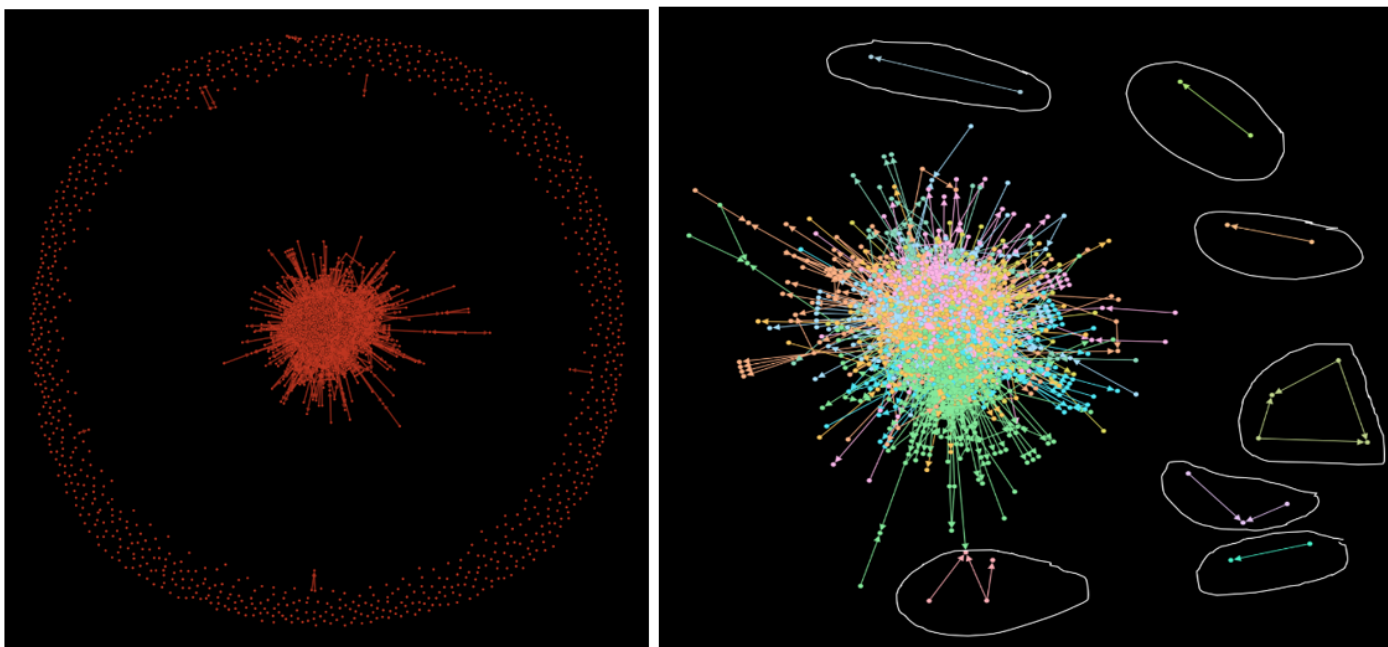


Figure 1: Methodology of this study. Step-1 (&5): Data collection.





**Figure 3:** (left) citation network structure related to ChatGPT, (right)  $k=1$  core of the citation network with 15 clusters/communities (shown in different colours) with emerging clusters highlighted.

the Paper-FoR affiliation network. The creation of the Paper-FoR affiliation network is done using code developed by the authors. From the  $WF$  network, the co-affiliation network of FoRs denoted by  $FF$  can be formed as:

$$FF = WF^T \times WF \quad (1)$$

Where  $T$  represents the transpose of the matrix representation of the Paper-FoR affiliation network so that  $WF^T = FW$ .

Creation of citation network is also done in this step using the code developed by us (on data collected during first phase as well as second phase). The speciality of the new code is that it is capable of having publication identification numbers (pub. ids) available in Dimensions database as vertices. The limitation of VOSviewer that represents a document in the 'author-year format' is overcome by this. Our code thus, facilitates of ease of analysis in both phases and thereby making tracking of progress of subnetworks of our interest easier.

### Step-3: Science mapping analysis using network techniques

Once the  $FF$  network is created, it can be analyzed using popular network analysis techniques such as degree (number of links associated with a vertex), eigenvector (quality of connections), etc., Formal (brief) definitions of these are given below:

Degree of a vertex  $i$  of a network (having  $n$  vertices in total) with adjacency matrix  $A = (a_{ij})$  is given by:

$$deg_i = \sum_{j=1}^n a_{ij} \quad (2)$$

Eigenvector score of a vertex  $i$  of a network with adjacency matrix  $A = (a_{ij})$  is given by:

$$eig_i = \frac{1}{\lambda} \sum_{j=1}^n a_{ij} x_j \quad (3)$$

Where  $\lambda$  is the largest eigenvalue of the matrix  $A$  and  $x_j$  represents the centrality score of neighbor  $j$ .<sup>[17]</sup> Hence, for a higher eigenvector, a vertex is required to have highly connected neighbours (reflects quality of connections) rather than more neighbours (direct connections), making it a relative centrality score that lies between  $[0,1]$ . The closer the score of a vertex to 1, the more quality/importance attributable to that vertex.

These will help to identify key FoRs associated with ChatGPT according to the quantity and quality of connections. Network clusters (Figure 2) are determined using VOS clustering,<sup>[18]</sup> and related analysis is also done.

### Step-4 (& 5): Citation network mining using the FV model and cluster analysis

This step is also carried out in two phases marked by  $t$  and  $t+T$ , but in different manners as described as follows.

For citation network analysis at  $t$ , the first task is the elimination of isolated vertices (disconnected vertices). This is done using a  $k$ -core filter,<sup>[19]</sup> with  $k=1$ .<sup>[15]</sup> The concept of  $k$ -core is briefly discussed below.

A  $k$ -core or core of the order  $k$  is a subnetwork of the original network such that all the vertices in the set of vertices induced by the subnetwork will have a degree (number of links) of at least  $k$ . According to this definition, zeroth order core or  $k=0$  represents the original network itself (including isolated or disconnected

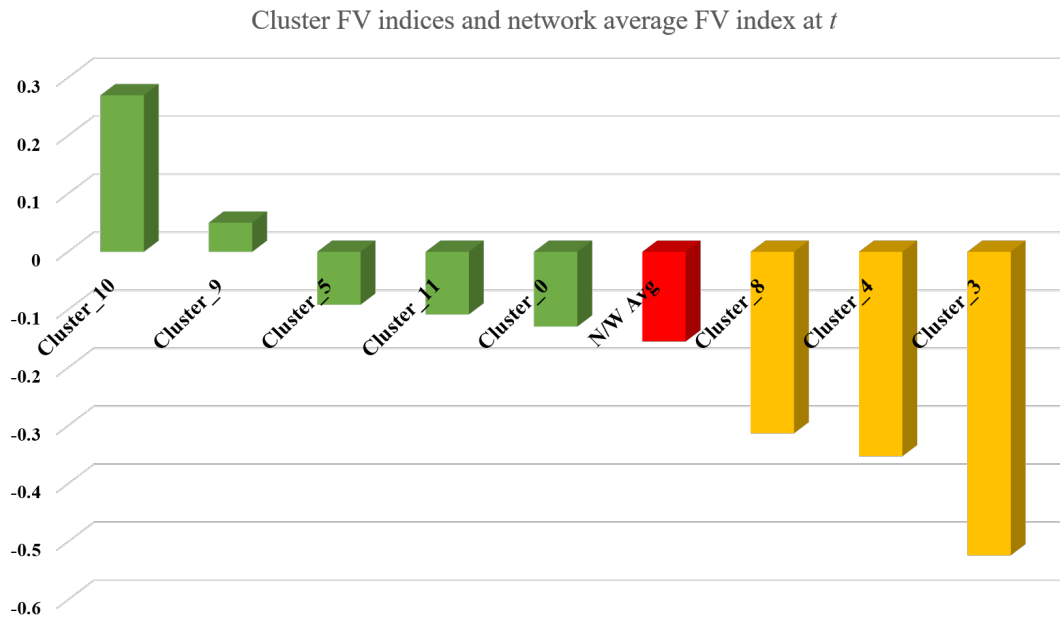


Figure 4: FV indices of grown clusters and the network FV index.

vertices) while core of the order 1 or  $k=1$  gives the subnetwork of the network with connected vertices only (after eliminating the isolates or disconnected vertices).

Cluster formation of the  $k=1$  core of the citation network is the next step. Prabhakaran *et al.*,<sup>[15]</sup> used the Line Island Formation algorithm,<sup>[20]</sup> to determine clusters that reveal different naturally formed connected components within the literature analyzed. However, the analyzed topic in that study was related to the field of ‘Information Technology’ for the research area ‘Engineering’, which was a well-developed topic, and hence enough components (apart from the giant component) were there for gathering insights. In this case, ChatGPT is a novel topic, but witnessing very fast growth. Thus, significant components other than the giant component are difficult to find. Thus, an algorithm that partitions the giant component fairly and other components (if it is possible) is required for early scientometric analysis. Thus, a fast community detection algorithm by Blondel *et al.*,<sup>[21]</sup> is used in this work. A brief description of the algorithm is given below.

### Fast Community Detection Algorithm

Communities or sub-units are sets of highly interconnected vertices, the determination of which is crucial for retrieving useful information from the large network structure. The fast community detection algorithm is designed to optimize modularity and assigns each vertex to different partitions initially and then proceeds by computing the modularity gain by placing vertex  $i$  from its community to the community of vertex  $j$ . The gain in modularity for a community for moving a vertex  $i$  to a community  $C$  can be computed using:

$$\Delta Q = \left[ \frac{\sum in+ki,in}{2m} - \left( \frac{\sum tot+ki}{2m} \right)^2 \right] - \left[ \frac{\sum in}{2m} - \left( \frac{\sum tot}{2m} \right)^2 - \left( \frac{ki}{2m} \right)^2 \right] \quad (4)$$

where  $\sum in$  is the sum of the weights of the links inside community  $C$ ,  $ki$ ,  $in$  is the sum of the weights of the links from  $i$  to vertices in  $C$ ,  $m$  is the sum of the weights of all the links in the network,  $\sum tot$  is the sum of the weights of the links incident to vertices in  $C$ ,  $ki$  is the sum of the weights of the links incident to node  $i$ . The second phase of the algorithm creates a new network by treating communities in phase 1 as vertices and links between the communities as links (weights computed as the sum of weights of links between vertices of the corresponding communities). Upon completion of the second phase, the first phase is reapplied on the new network to iterate (for optimization of modularity) until no more change is possible and maximum modularity is attained. The advantages of this algorithm include speed, ease of implementation and freedom from resolution limit problem, making it the best choice.

Thus, once community (cluster) determination is done, substantially grown and emerging clusters can be identified. However, for determination of relative importance of clusters in terms of state of knowledge flow and thereby determining key clusters, FV model has to be applied.

Now, for applying the FV model, the FV indices of each publication need to be computed. A brief description of the FV model is attempted.

*FV model:* Being an information network, the citation network of publications is supposed to have a knowledge/information flow from cited papers to citing papers. There are four possible types of works in a network, viz. (i) papers that received knowledge from other works, but have not transmitted knowledge to other works, (ii) papers that have not received knowledge from other works, but have transmitted knowledge to other works, (iii) works that

have received knowledge from as well as transmitted knowledge to other works, and (iv) works that have neither received nor transmitted knowledge to other works. The works in the fourth category are isolated and are found to be disconnected from the network and the merit of those cannot be determined from network analysis. All the other three categories are supposed to have a flow vergence (i.e., convergence or divergence) potential that can be estimated using network analysis methods.

The *flow vergence potential* of a work is the ability of a work to transmit knowledge/information (directly as well as indirectly) to other works and thereby contribute to the formation of knowledge pathways and also to the growth of the cluster (specific sub-body of literature) and also to the whole network (body of literature) to which it belongs to. Network concepts that can be used to reflect the flow vergence potential are discussed next. A work is supposed to be in flow divergence mode if its flow vergence potential is high ( $> 0$ ) and flow convergence mode if its flow vergence potential is less than 0. High flow divergence potential means work is either already having a higher outflux of knowledge (reflected by indegree) than influx (reflected by outdegree) and will continue or might soon have a higher outflux.

Using these basic measures of direct knowledge flow and indirect knowledge flow (reflected by eigenvector) the flow vergence potential can be expressed as:

$$W_{FV_i} = \frac{\text{indeg}_i - \text{outdeg}_i}{\text{indeg}_i + \text{outdeg}_i} + \text{eig}_i \quad (5)$$

The centrality measures *indeg* and *outdeg* are local measures, while *eig* is a global measure, that reflects indirect knowledge transfer too (and also reflects the quality of knowledge transfer) making FV index a hybrid centrality measure.

A work in 'flow divergence potential' can contribute more to the growth of the network or the cluster to which it belongs, as it may get cited by other works or the works that cited it may get citations where such a possibility may be less for a work-in-flow convergence mode unless it attains flow divergence mode.<sup>[15]</sup> The theoretical and rational aspects of the FV model are discussed in detail by Prabhakaran *et al.*,<sup>[15,22,23]</sup> and Lathabai *et al.*<sup>[24,25]</sup>

Now, the network average FV index represents a network's overall flow vergence mode. If it is greater than 0, the network is dominated by works in flow divergence mode (the presence of so many works in flow divergence mode to ensures considerable growth of the network), and if otherwise, more works need to become in flow divergence mode to ensure considerable growth. The network average FV index can be computed as given by Prabhakaran *et al.*<sup>[15]</sup>

$$N_{FV} = \frac{\sum_i^M W_{FV_i}}{M} \quad (6)$$

Where  $M$  is the total number of publications in the network.

Similarly, the cluster average FV index represents a cluster's overall flow vergence mode. If it is greater than 0, the presence of

works in flow divergence mode in the cluster is enough to ensure its substantial growth otherwise, the cluster requires more works in flow divergence mode. The cluster average FV index can be computed as given by Prabhakaran *et al.*<sup>[15]</sup>

$$Clus_{FV} = \frac{\sum_i^K W_{FV_i}}{K} \quad (7)$$

Where  $K$  is the total number of publications in the cluster.

Now according to the FV model, if  $> N_{FV}$ , such clusters are more contributive to the network than the rest of the clusters (which, having  $< N_{FV}$ ).

Determining the themes of the clusters formed using the community detection algorithm is very difficult as it requires content analysis of all the publications. In the case of ChatGPT, since most of the works may be dealing with applications of ChatGPT in various fields, there are chances for finding works dealing with different applications (themes) within the same cluster. Therefore, the term/theme that appears in most of the publications in a cluster is used to represent the cluster. This can be done by using the *WF* network. Firstly, the *WF* network must be partitioned with the  $C_i$  ( $k=1$ ) cluster membership information using a matching and filtration process. Then, for each cluster, the *indeg* of FoRs needs to be found, which will give the number of publications to which that FoR is tagged. The FoR with the highest *indeg* can be assigned as the dominant theme of the cluster. The top papers with the highest FV indices can determine key research papers within each cluster. Content analysis of these will help to extricate useful insights for the stakeholders.

For second phase, almost all the processes are repeated for citation network created at  $t+T$ . Cluster formation, network FV index computation and cluster FV indices computation is done at  $t+T$  in the same way as done for network at  $t$ . However, as analysis at this phase is intended to track the growth of grown and emerging clusters identified at  $t$ , determination of themes of the clusters found in new network, content analysis of selected publications, etc., are not done.

Insights obtained from analyses at steps 3, 4 and 5 can address objective 6.

## RESULTS AND DISCUSSION

After data collection (at first phase), the next step is the network creation of *FF* and *Ci* networks. *FF* network (with clusters) is given in Figure 2.

The *FF* network is undirected (links are not directed), consisting of 155 vertices (FoRs) and 698 links between them. Upon degree analysis (see Table 1), the most connected FoR in the science map is '46 Information and Computing Sciences', followed by '32 Biomedical & Clinical Sciences'. Out of the top 5, three FoRs occupy the next three FoRs- '40 Engineering', '47 Language, Communication & Culture', and '35 Commerce, Management, Tourism & Services', occupies the next three positions with the

same connectivity. Here, the 2-digit prefix numbers (like 46, 47, 32, etc.,) are identifiers provided by ANZRC to the major FoRs (top level in the hierarchy named as 'Division') and 4-digit prefix numbers (like 4608, 3211, etc.,) represent sub FoRs (level just below top level in the hierarchy named as 'Group'). ChatGPT-related research in FoR '46 Information and Computing Sciences' is directly influencing 84 fields, and it is interesting to note that ChatGPT-related research in '32 Biomedical & Clinical Sciences' occupies the next position. As fields '40 Engineering' and '35 Commerce, Management, Tourism & Services' are equally impacting many fields, several industries might be going through technological and business model innovations impacted by ChatGPT-related research in these FoRs. The rise of ChatGPT also provided the FoR '47 Language, Communication & Culture' a good position in a 'science map', which does not happen that often. Apart from '46 Information and Computing Sciences', the FoR '47 Language, Communication & Culture' can directly influence ChatGPT evolution too. This may be the reason for the higher connectivity of this field in the science map.

Interestingly, by virtue of the quality of connections, '47 Language, Communication & Culture' occupies a high position (second

**Table 1: Top 5 FoRs according to degree analysis of the FoR network.**

Fields of Research	Degree
46 Information & Computing Sciences	84
32 Biomedical & Clinical Sciences	50
40 Engineering	39
47 Language, Communication & Culture	39
35 Commerce, Management, Tourism & Services	39

**Table 2: Top 5 FoRs according to Eigenvector analysis of FoR network.**

Fields of Research	Eigenvector
46 Information & Computing Sciences	1
47 Language, Communication & Culture	0.654
32 Biomedical & Clinical Sciences	0.587
35 Commerce, Management, Tourism & Services	0.581
4608 Human centered Computing	0.565

**Table 3: Cluster FV indices of the grown clusters and their themes.**

Cluster id.	FV index	Status	Theme
Cluster_10	0.2698	Divergence	46 Information & Computing Sciences
Cluster_9	0.0502	Divergence	32 Biomedical & Clinical Sciences
Cluster_5	-0.0912	Convergence	32 Biomedical & Clinical Sciences
Cluster_11	-0.1079	Convergence	39 Education
Cluster_0	-0.1185	Convergence	40 Engineering
Cluster_8	-0.3129	Convergence	32 Biomedical & Clinical Sciences
Cluster_4	-0.3524	Convergence	32 Biomedical & Clinical Sciences
Cluster_3	-0.5232	Convergence	32 Biomedical & Clinical Sciences

position in Table 2). Though '32 Biomedical & Clinical Sciences' has high direct connectivity, the FoRs that are directly mapped to the FoRs connected with '32 Biomedical & Clinical Sciences' have relatively less connectivity to other fields. This can indicate the following. Though the ChatGPT-related research happening in '32 Biomedical & Clinical Sciences' is having a direct impact on several fields, the fields that are impacted are not having much impact on many other fields and hence the possibility of a 'transitive impact' like '46 Information and Computing Sciences' and '47 Language, Communication & Culture' is not there.

'4608 Human-centered computing', a subcategory of '46 Information and computing Sciences', is also found to be having good eigenvector score, and hence it can be regarded as the most qualitatively influencing subfield of the FoR '46 Information and computing Sciences'. This subfield is specifically important in the current and coming times as several concerns are already associated with human safety and many other aspects of the growth of AI in general and generative AI. Now, the citation network analysis results using the FV model can be examined.

The citation network (a directed network) created from collected data is shown in Figure 3 (left). It consists of 1873 papers and 3145 citation links. So many disconnected papers (neither cited by nor cited any other papers) were removed by applying the  $k=1$  core filter. Our further analyses use the resultant network with 858 publications and 3145 links shown in Figure 3 (right). The network has 7 connected components (1 giant and 6 small). For cluster analysis, cluster determination of the resultant network is done using a community detection algorithm, and 15 clusters are obtained (with id.s ranging from 0 to 14). The  $k=1$  core of the citation network, along with clusters (in different colours), is shown in Figure 3 (right). Out of the 15 clusters (communities), 8 are grown clusters (found in the giant component), and 7 are emerging clusters (one cluster, i.e., Cluster\_7, is found inside the giant component and other 6 are the small components of the network). Within a short span, ChatGPT-related literature is exhibiting the behaviour of a substantially grown topic (with a giant and some small components and also showing many grown and emerging clusters).

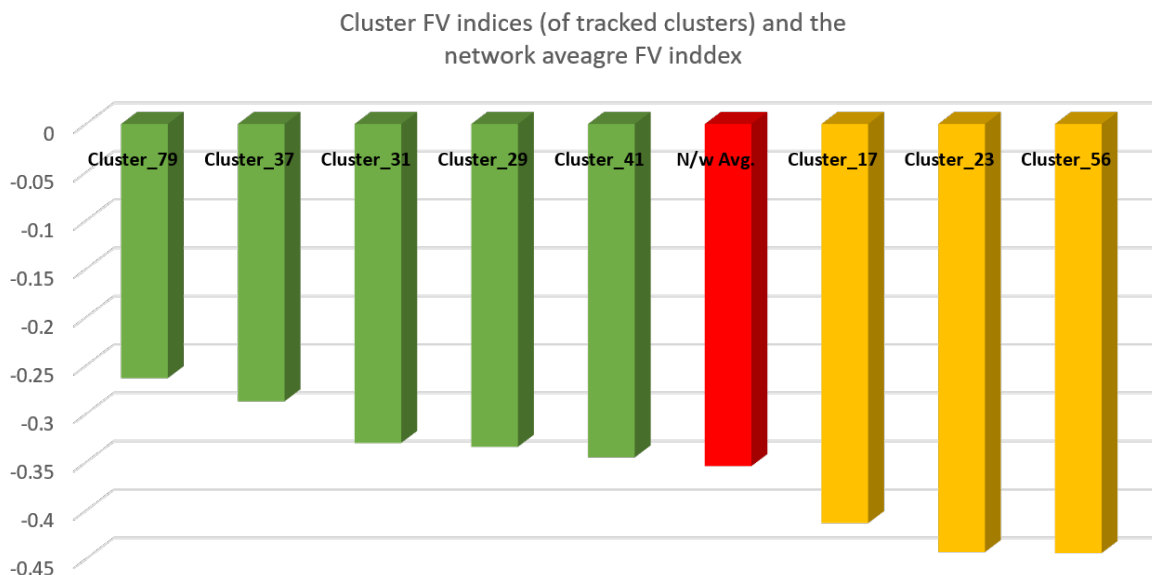


Figure 5: Cluster FV indices of 8 clusters formed at  $t+7$  out of 15 clusters at  $t$ .

Table 4: Themes of the emerging clusters.

Cluster id.	Theme
Cluster_1	46 Information & Computing Sciences
Cluster_2	32 Biomedical & Clinical Sciences
Cluster_6	46 Information & Computing Sciences
Cluster_7	46 Information & Computing Sciences
Cluster_12	50 Philosophy & Religious Studies
Cluster_13	46 Information & Computing Sciences
Cluster_14	39 Education

Table 5: Top specific contribution details of Clusters 10 and 9.

Cluster id.	Publication id.	Title of publications
Cluster_10	pub.1154987241	ChatGPT: Open Possibilities.
	pub.1155471102	What Can ChatGPT Do? Analyzing Early Reactions to the Innovative AI Chatbot on Twitter.
Cluster_9	pub.1155855793	What Does ChatGPT Say: The DAO from Algorithmic Intelligence to Linguistic Intelligence.
	pub.1154639234	Artificial Intelligence Discusses the Role of Artificial Intelligence in Translational Medicine A JACC: Basic to Translational Science Interview with ChatGPT.
	pub.1156014992	Using ChatGPT to write patient clinic letters.
	pub.1156413712	Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma.

After the computation of FV indices of publications using (5), the network FV index and cluster FV indices are found using (6) and (7), respectively. The network FV index is found to be  $-0.1546 (< 0)$ , and hence the network is in convergence mode, which means that more publications with divergence potential are required to ensure the growth of the network (literature) at a higher pace. The cluster FV indices of the grown clusters and the Network average FV index are shown in Figure 4.

Clusters 10 and 9 are in divergence mode (and are high performing), while clusters 5, 11, 0, 8, 4, and 3 are in convergence mode. However, Clusters 10, 9, 5, 11, and 0 are contributing more to the growth of the network (shown in green colour in Figure 4.) as these clusters' average is higher than the network average (red colour). Clusters 8, 4, and 3 are relatively less performing and are shown in yellow in Figure 4. The themes of clusters are determined using the method specified in section 2 with the help of the *WF* network. The themes of the 8 grown clusters and FV indices values are given in Table 3. It is interesting to note the presence of '32 biomedical and clinical sciences' as a dominant

theme in most grown clusters. Themes '40 Engineering' & '39 Education' dominated two grown clusters. This achieves the second objective.

For achieving the third objective, the themes of emerging clusters need to be identified. These are also done using the procedure mentioned in section 2 with the help of the *WF* network.

**Table 6: Top specific contribution details of clusters 5 and 11.**

Cluster id.	Publication id.	Title of publications
Cluster_5	pub.1153838233	Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models.
	pub.1155156738	Generating scholarly content with ChatGPT: ethical challenges for medical publishing.
Cluster_11	pub.1155156740	ChatGPT: friend or foe?
	pub.1153518825	AI bot ChatGPT writes smart essays. should professors worry?
	pub.1154007278	Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers.
	pub.1154069231	OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare.

The themes determined for the emerging clusters are given in Table 4. These clusters are too small to discuss flow vergence status. Most emerging clusters have '46 Information and Computing Sciences' as the dominant theme. Themes '32 Biomedical & Clinical Sciences', '39 Education' and '50 Philosophy & Religious Studies' dominated three clusters.

Now, for determining key specific research contributions within grown and emerging clusters, content analysis of the top three publications with the highest FV index values and content analysis of all the publications in emerging clusters are done. The details of the same are discussed next. Details of clusters 10 & 9 are given in Table 5, clusters 5 & 11 are given in Table 6, clusters 0 & 8 in Table 7, and clusters 4 & 3 in Table 8.

**Cluster\_10:** The top-most paper in this cluster enlisted and discussed the role of ChatGPT can play in academic writing, search engines, coding, vulnerability detection and social media.<sup>[26]</sup> The second work analyzed early tweets (the first month after the launch of ChatGPT) using LDA and identified functional domains that can be impacted like creative writing, essay writing, prompt writing, code writing, and answering questions.<sup>[27]</sup> The third work observed limitations of ChatGPT to revolutionize large scale research and linguistic intelligence, but took an optimistic stance on its potential to contribute to revolutionize these and Industry 5.0.<sup>[28]</sup>

**Table 7: Top specific contributions of Clusters 0 and 8.**

Cluster id.	Publication id.	Title of publications
Cluster_0	pub.1156217344	Role of Chat GPT in Public Health.
	pub.1155847243	Potential Use of Chat GPT in Global Warming.
Cluster_8	pub.1156180076	Will ChatGPT transform healthcare?
	pub.1155222253	How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment.
	pub.1155380932	Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study.
	pub.1154371389	A conversation with ChatGPT on the role of computational systems biology in stem cell research.

**Table 8: Top specific contributions of clusters 4 and 3.**

Cluster id.	Publication id.	Title of publications
Cluster_4	pub.1154888867	ChatGPT is fun, but not an author.
	pub.1153880110	Rapamycin in the context of Pascal's Wager: generative pre-trained transformer perspective.
Cluster_3	pub.1154432832	Abstracts written by ChatGPT fool scientists.
	pub.1154632837	ChatGPT listed as author on research papers: many scientists disapprove.
	pub.1154707004	The AI writing on the wall.
	pub.1155149354	What ChatGPT and generative AI mean for science.

**Cluster\_9:** In the topmost paper,<sup>[29]</sup> the role of AI in translational medicine was prompted to ChatGPT and found that though responses lacked nuanced insight expected from a person with in-depth knowledge, indicating it does not have to go a long way to become an important voice in scientific and medical journals. The second paper evaluated the readability, factual correctness, and humanness of ChatGPT-generated clinical letters to skin cancer patients and emphasized the need for caution and proactive addressal of potential risks should be ensured for the safety and quality of patient care.<sup>[30]</sup> The third work examined the accuracy and reproducibility of ChatGPT in answering questions related to knowledge, management, and emotional support for cirrhosis and HCC, concluded that ChatGPT can have an adjunct role as an informational tool for patients and doctors to improve outcomes.<sup>[31]</sup>

**Cluster\_5:** The topmost paper evaluated the performance of ChatGPT on three staged US medical licensing exams, in which ChatGPT performed at or near passing threshold in all three steps, suggestive of its potential to be used for medical education as well as clinical decision-making.<sup>[32]</sup> The second work argued that the possibility of generating inaccurate or misleading text by ChatGPT strengthens the concern of scholarly misinformation and emphasized the need for comprehensive guidance for AI-generated content usage within scholarly publishing at the earliest.<sup>[33]</sup> The third work opined that with the advancement of AI generative technologies, editorial policies need to evolve too, directing more forethought, oversight and investment towards model training, and AI output detectors as humanity is not yet ready for a 'game change'.<sup>[34]</sup>

**Cluster\_11:** The topmost work in this '39 Education' dominated cluster expressed concerns on potential effect on human knowledge and ability (especially 'thinking') and stresses the responsibility of academia for imposing a healthy distrust to properly deal with this insurmountable challenge.<sup>[35]</sup> The second work evaluated the abstracts generated by ChatGPT using an Artificial Intelligence (AI) output detector and plagiarism detector and blinded human reviewers and found that human reviewers detected 68% of generated abstracts correctly but incorrectly identified 14% of original abstracts as generated, pointing towards the gravity of challenges in review and assessment of scientific articles.<sup>[36]</sup> The third work explored the capability of ChatGPT to create a literature review article on the topic 'digital twin in healthcare' and concluded that the future academic publishing process will require less human effort, which in turn will allow academics to focus on their research.<sup>[37]</sup>

**Cluster\_0:** The topmost paper observed that ChatGPT has the potential to provide informed decisions about health for individuals and communities and enlisted its limitations and challenges, advantages and disadvantages too.<sup>[38]</sup> The second one discussed how AI and NLP technologies, such as ChatGPT, can

play a crucial role in understanding climate change and improving the accuracy of climate projections, especially for model parameterization, data analysis and interpretation, scenario generation, and model evaluation.<sup>[39]</sup> The third paper stated that ChatGPT can massively transform healthcare if specific clinical needs are identified for its application.<sup>[40]</sup>

**Cluster\_8:** The topmost work in this cluster evaluated and analyzed the performance of ChatGPT on the United States Medical Licensing Examination (USMLE) and found that the model is comparable to third-year medical students, strengthening the possibility of usage of ChatGPT as a medical education tool.<sup>[41]</sup> The second work with the highest FV index Compared the knowledge and interpretation ability of ChatGPT with medical students in parasitology and concluded that ChatGPT's knowledge and interpretation ability for this parasitology examination was not yet comparable to those of medical students in Korea.<sup>[42]</sup> The third work explored with the help of ChatGPT, how advances in computation can help practitioners by saving time and freeing researchers (computational systems biology for example) and found that many of the responses lacked depth and insight.<sup>[43]</sup>

**Cluster\_4:** The first work in this cluster emphasized the need to update license and editorial policies to undertake actions (similar to scientific misconduct, image alteration and plagiarism) against usage of not only text, but also figures, images, or graphics generated by ChatGPT while excluding legitimate AI generated dataset.<sup>[44]</sup> The second work explored the benefits of Rapamycin usage in the context of the philosophical argument 'Of Pascal's Wager' with the help of ChatGPT, which successfully picked up its pros and cons and provided wise recommendations to consult health care professionals.<sup>[45]</sup> The third one outlined the possibility of misleading researchers and society through flawed research conducted with ChatGPT's help and observed that more or equal focus should be on incentives that lead to publication pressure that force researchers to go for desperate measures.<sup>[46]</sup>

**Cluster\_3:** The topmost work in this cluster discussed the tendency of some authors to list ChatGPT as an author in the byline information, and as ChatGPT cannot kept responsible or accountable, forces editors and publishers are forced to devise suitable policies to restrict the use of ChatGPT.<sup>[47]</sup> The second one emphasized the need and standardization (as in the case of plagiarism services) of applications like GPTZero to detect text generated by ChatGPT.<sup>[48]</sup> The third work discussed various aspects of generative AI reviewed the technology's development and usage status and opined that with judicious oversight from experts, LLMs may even aid cancer diagnosis by cross-checking body scan images and academic literature.<sup>[49]</sup>

Now, seven emerging clusters are analyzed. Firstly, clusters related to '46 Information & Computing Sciences' are analyzed. Details of publications in clusters 1 and 13 (two small clusters) are given

in Table 9. Details of clusters 6 and 7 are given in Tables 10 and 11, respectively.

**Cluster\_1:** The first work given in Table 9 in this cluster observed that ChatGPT's moral advice is inconsistent and emphasized on improvement of Users' digital literacy along with the better design of ChatGPT and other tools to remain immune from corruption.<sup>[50]</sup> The other work introduced logic-symbolic inference that has the capability to deal with various belief

**Table 9: Specific contribution details of clusters 1 and 13.**

Cluster id.	Publication id.	Title of publications
Cluster_1	pub.1157067307	ChatGPT's inconsistent moral advice influences users' judgment.
	pub.1157571557	Epistemic considerations when AI answers questions for us.
Cluster_13	pub.1156408128	Artificial muses: Generative Artificial Intelligence Chatbots Have Risen to Human-Level Creativity.
	pub.1155273939	The Artificial Creatives: The Rise of Combinatorial Creativity from Dall-E to GPT-3.

**Table 10: Specific contribution details of Cluster 6.**

Cluster id.	Publication id.	Title of publications
Cluster_6	pub.1156582292	How Trustworthy is ChatGPT? The Case of Bibliometric Analyses.
	pub.1156501765	ChatGPT as a Tool for Bibliometrics Analysis: Interview with ChatGPT.
	pub.1157011331 & pub.1157261271	Network Visualization of ChatGPT Research: a study based on term and keyword co-occurrence network analysis.

**Table 11: Specific contribution details of Cluster 7.**

Cluster id.	Publication id.	Title of publications
Cluster_7	pub.1157001443	Using ChatGPT to Fight Misinformation: ChatGPT Nails 72% of 12,000 Verified Claims.
	pub.1157452004	Testing the Reliability of ChatGPT for Text Annotation and Classification: A Cautionary Remark.
	pub.1157451552	"HOT" ChatGPT: The promise of ChatGPT in detecting and discriminating hateful, offensive, and toxic comments on social media.
	pub.1157618732	Is ChatGPT better than Human Annotators? Potential and Limitations of ChatGPT in Explaining Implicit Hate Speech.

systems to handle any possible epistemic associated with human or artificial information processors as reliance on AI to answer our questions and judge our output raises some epistemic concerns.<sup>[51]</sup>

**Cluster\_13:** The first work in this cluster given in Table 9 conducted a comparative analysis of human-generated ideas and six GAIs, including ChatGPT and registered no qualitative difference, indicating GAI's potential as creative assistants (upon further technological improvement).<sup>[52]</sup> The other work considered and compared the impacts of generative AI such as Dall-E, Jasper, or ChatGPT-3 on artists and creatives and provided suggestions to minimize the likely disruption to the creative market.<sup>[53]</sup>

**Cluster\_6:** The first work in Table 10 evaluated ChatGPT's performance on bibliometric analysis by comparing an existing bibliometric study on a topic with output provided by the chatbot and directed researchers to exercise caution when using ChatGPT as a bibliometric tool.<sup>[54]</sup> The second work revealed the benefits and problems of ChatGPT usage as a tool for bibliometric analysis by prompting three valid questions to ChatGPT.<sup>[55]</sup> The third work and fourth works (preprint versions in different archives) identified the major research areas of ChatGPT through term and keyword co-occurrence network mapping techniques, and revealed key terms like AI, LLMs and GPT.<sup>[56]</sup>

**Cluster\_7:** The first work in Table 11 explored the ability of LLMs like ChatGPT to remedy the surge of online misinformation, and demonstrated the potential of ChatGPT to improve content moderation practices and complement the work of human fact-checking experts.<sup>[57]</sup> The second work the reliability of ChatGPT for text annotation and classification cautioned researchers to use ChatGPT for zero-shot test annotation, underscoring the necessity for thorough validation against human-annotated data.<sup>[58]</sup>

The third work investigated the potential of ChatGPT to understand and detect harmful content by comparing its performance with MTurker annotations for concepts related to harmful content such as Hateful, Offensive, and Toxic (HOT) and provided several insights about the reliability and consistency of ChatGPT in detecting harmful content in social media.<sup>[59]</sup> The

fourth work examined ChatGPT's ability to provide Natural Language Explanations (NLEs) for implicit hate speech detection and laid out its potential and limitations for the same.<sup>[60]</sup>

Now the details of the remaining emerging Clusters 2, 12 and 14 are given in Table 12.

**Cluster\_2:** The first paper in this cluster, given in Table 12, aimed to determine if ChatGPT could be utilized to produce novel systematic review ideas related to Plastic Surgery and found that it may be useful for virtual consultations, pre-operative planning, patient education, and post-operative patient care.<sup>[61]</sup> The second paper opined that although ChatGPT may remedy difficult issues in plastic surgery, a new design of ChatGPT, examination and modification of ChatGPT code of conduct is required for applying it for practice, study or instruction, research, etc.,<sup>[62]</sup>

**Cluster\_12:** The first work in this cluster given in Table 12 prompted ChatGPT about how history would have changed had the attempted assassination of Franklin D. Roosevelt succeeded and the replies strengthened the conjecture that ChatGPT has been programmed to have a 'left wing' political bias.<sup>[63]</sup> The second work in this cluster prompted ChatGPT on whether Environmental, Social, and Governance (ESG) is a bad idea and upon a series of follow-up queries, a 'left-wing' political bias

**Table 12: Specific contribution details of Clusters 2, 12 and 14.**

Cluster id.	Publication id.	Title of publications
Cluster_2	pub.1156378863	Utilization of ChatGPT for Plastic Surgery Research: Friend or Foe?
	pub.1157735882	Utilization of ChatGPT for Plastic Surgery Research: Comment.
Cluster_12	pub.1157512963	How Would American History Be Different If FDR Had Been Assassinated in 1933: A Chatgpt Essay.
	pub.1157510581	Is ESG a Bad Idea? The Chatgpt Response.
	pub.1157527348	How to Determine Your 'Fair Share' of Taxes: Ask Chatgpt.
Cluster_14	pub.1157966422	Can ChatGPT and Bard generate aligned assessment items? A reliability analysis against human performance.
	pub.1158091256	Guidance for researchers and peer-reviewers on the ethical use of Large Language Models (LLMs) in scientific research workflows.

behind the coders of ChatGPT is revealed.<sup>[64]</sup> The third work prompted ChatGPT about the taxation systems and it provided a balanced answer.<sup>[65]</sup>

**Cluster\_14:** The first paper in this cluster, given in Table 12, examined the reliability of OpenAI ChatGPT and Google Bard LLM tools against experienced and trained humans and found that ChatGPT and Google Bard's inter-reliability was low against gold standard human ratings.<sup>[66]</sup> The second work attempted to provide guidance and norms for applications of LLMs and for peer review of LLM research and opined that by ensuring alignment of LLM research with AI ethics, the use of LLMs with ethical principles and best practices can be ensured.<sup>[67]</sup>

With this, the fourth objective of this research is achieved. The fifth objective, i.e., tracking of the growth of the identified (in phase 1) grown and emerging clusters. This is discussed next.

### Tracking of growth of grown and emerging clusters

By December 31, 2024 (phase 2 of our analysis), total publications have increased up to 33046. Upon the application of the  $k$ -core filter ( $k=1$ ), the resultant network had 11627 papers and 45319 links. This network had 250 connected components. There were 284 communities/clusters. The giant component itself hoards 11027 publications belonging to 35 different clusters/communities (earlier 8 grown clusters and 1 emerging cluster were present). Visualization of this network and its communities is not possible due to its sheer size.

Now, upon analysis, it is found that almost all the grown clusters have grown considerably, and so do emerging clusters. A more intriguing fact is that some of the grown clusters are found to merge (citations received from new entrants in the network glued those together) to form much larger clusters. More information on this can be obtained from Table 13.

From Table 13, it is clear that 8 clusters are formed at  $t+T$  out of 15 clusters present at  $t$  with or without merger and all of them are substantially grown. Hence, all these 8 clusters are eligible for cluster FV computation. Cluster FV indices and network FV index computed for 8 clusters are shown in Figure 5. The Network FV index is found to have decreased to -0.354 (from -0.1546 at  $t$ ), indicating that network is in convergence mode (in a greater extent than that at  $t$ ). This may be due to the flooding of so many new entrant publications into the network, that clearly outnumbered and outweighed the works that might have increased its FV potential and switched to flow divergence mode from convergence mode. However, abrupt growth indicates that works identified as the ones with high divergence potential might have garnered direct connectivity (citations) or indirect connectivity and played a crucial role in knowledge flow (as stipulated by the FV model). This exercise, in a sense, can be treated as a validation of FV model.

**Table 13: Cluster growth found on the second phase.**

May 2023 (Phase 1)			December 2023 (Phase 2)		
Cluster id.	Growth type	# of papers	Cluster id.	Merger status	# of papers
Cluster_4	Grown	152	Cluster_17	Merger of clusters 4, 3, 5, 11 & 6	1759
Cluster_3	Grown	63	Cluster_17		
Cluster_5	Grown	115	Cluster_17		
Cluster_11	Grown	177	Cluster_17		
Cluster_6	Emerging	4	Cluster_17		
Cluster_10	Grown	70	Cluster_29	Merger of clusters 10 & 0	1565
Cluster_0	Grown	41	Cluster_29		
Cluster_8	Grown	112	Cluster_23	Merger of clusters 8 & 9	2253
Cluster_9	Grown	109	Cluster_23		
Cluster_1	Emerging	2	Cluster_37	Nil	617
Cluster_13	Emerging	2	Cluster_41	Merger of clusters 13 & 12	712
Cluster_12	Emerging	3	Cluster_41		
Cluster_7	Emerging	4	Cluster_31	Nil	600
Cluster_2	Emerging	2	Cluster_56	Nil	264
Cluster_14	Emerging	2	Cluster_79	Nil	455

Coming to tracking of growth of grown clusters, Table 13 clearly indicates that cluster \_17 is formed out of the merger of 4 grown clusters (clusters 3, 4, 5 and 11) and one emerging cluster (cluster\_6). Thus, the merger of two relatively better performing clusters (clusters 5 and 11), two lesser performing clusters (clusters 3 and 4) and an emerging cluster (whose performance was not possible to be assessed at  $t$ ) resulted in the formation of a relatively less performing cluster (at  $t+T$ ). However, from Figure 5, cluster\_17 is the best among lesser performing clusters (might have been mostly due to the effect of publications with high FV potential from clusters 5 and 11). The highest performing cluster at  $t$  (cluster\_10) and another good performing cluster (cluster\_0) merged to form cluster\_29 at  $t+T$ , which is one of the better performing clusters (Figure 5), despite its large size at  $t+T$ . On the other hand, two large-sized clusters at  $t$ , cluster\_9 (second best cluster at  $t$ ) and cluster\_8 (a low performing cluster) have merged to form a huge cluster viz. cluster\_23, which is one of the lesser performing clusters at  $t+T$ .

It is interesting to note that most of the emerging clusters have grown and are found to be highly performing. Clusters 1, 7 and 14 are found to be grown without any merger. Cluster\_79 (new cluster id. of cluster\_14) is the best performing cluster at  $t+T$ , closely followed by clusters 37 and 31 (earlier clusters 1 and 7, respectively). Clusters 13 and 12 (at  $t$ ), merged to form cluster\_41, which is the least performing among the relatively high performing clusters at  $t+T$ . However, cluster\_56 (earlier cluster\_2) is the least performing in the whole group of 8 clusters,

which is the least grown cluster among the clusters formed out of emerging clusters, possibly indicating the relative less appeal of its theme.

Thus, the framework that supports dynamic analysis of the FV model is found to be highly useful for the analysis of themes that exhibit a gold rush (like in the case of 'ChatGPT'). The dynamic analysis has shown that the merger of early identified high performing grown clusters may result in relatively good performing clusters. Merger of high performing grown cluster with one or more lesser performing clusters might result in a relatively low performing cluster. Dynamic analysis framework can also help to determine promising emerging clusters. Thus, the fifth objective is also successfully addressed.

The sixth objective is to discuss insightful implications for various stakeholders. It is covered in the next section.

## Implications for various stakeholders

### *Manufacturers of ChatGPT and other Generative AI tools*

- From the very early FoR mapping analysis, '32 Biomedical & Clinical Sciences' and '47 Language, Communication & Culture' are the highest mapped FoRs quantitatively and qualitatively other than '46 Information & Computing Sciences', manufacturers of ChatGPT and other generative AI tools should continuously monitor the discussions from the field '32 Biomedical & Clinical

Sciences' to get updated information about the issues, biases related to ChatGPT and GAI, and challenges that should be addressed for the safest and most reliable application in the field. The discussions from '47 Language, Communication & Culture' can complement the technological developments in '46 Information & Computing Sciences' for improving the LLM backbone of the GAI.

- From the content analysis of the top works in different grown and emerging clusters, key implications include:
- The reliability of ChatGPT should be improved for its usage for assessment and peer review purposes, if possible, to achieve a gold standard by human experts.
- ChatGPT's ability should be improved greatly for medical advice to patients, practitioners, etc., as its capability to provide insightful recommendations is still not comparable to an expert in the field.
- ChatGPT and other GAI are not entitled to authorship (per most leading journals' policies). Some provisions should be considered to get informed about such attempts and resist them. For instance, if some researcher prompts ChatGPT to answer any open or novel research questions, either restrict from such attempt or report such attempts as per guidelines provided by legal or ethical committees of the respective countries or international bodies (if there are any).
- As the output from ChatGPT and GAI for literature review queries exhibit discrepancies with that of human-generated reviews, focus especially on improving the ability of GAI to generate clear, bias-less, and accurate reviews so that researchers could focus more time on conducting research.
- Either restrict ChatGPT from providing morality-related advice or train the GAI effectively with updated and wide set of norms and notions of morality prevailing in different regions so that answers should apply time, place, and other important factors that affect morality-related questions.
- As the potential to detect online misinformation and harmful content is crucial for ChatGPT and GAI, R&D activities can also be oriented towards that direction.
- GAI should be free from biases, especially political bias (as 'left wing' bias of coders is detected). Special measures should be there to ensure a balanced political perception of ChatGPT.

## International policymakers for education and research

From FoR mapping analysis, the top FoRs according to quantity and quality are revealed to be '46 Information & Computing Sciences', '32 Biomedical & Clinical Sciences', '47 Language, Communication & Culture', '40 Engineering', '39 Education' and '35 Commerce, Management, Tourism and Services', the following implications are there:

- As it is the right time to form policies using ChatGPT, other GAI, and other AI tools for research and education, form expert committees ensuring representation from all the above FoRs and nationalities for formulating robust policies.
- Special emphasis should be there on regulating the use of ChatGPT, other GAI, and other AI tools for education to ensure that it won't curb the ability of students to think and solve problems.
- Form special ethical committees to set ethical guidelines for using ChatGPT and other generative AI for education and research. Other than experts from the FoRs mentioned above, experts from the field '50 Philosophy and Religious Studies' should also be included.
- The formation of legal committees to form a legal framework in compliance with the ethical framework can also be done.
- Special emphasis should be made to encourage R&D in FoR '4608 Human-centered Computing' to ensure a research and education environment with safe, reliable, and human-friendly AI tools.
- From the analysis of top works in different grown and emerging clusters, the following implications are extracted:
- Regulations to tackle the authorship issues related to ChatGPT should be there. Regulations should aim for source-level reporting of attempts for unfair usage of GAI or AI tools for research by legally and ethically binding the manufacturers to restrict such uses or report such attempts to the respective committees. Also, the submitted (to Journals or other publication outlets) AI-generated text detected from AI output detectors can also be reported to ethical and legal committees.
- Ethical guidelines should specify that GAI and AI tool manufacturers should not be biased in favour of any political ideology and should be encouraged to maintain a balanced political perspective.
- As misinformation spreading through research publications is also common these days (as witnessed

during the COVID-19 pandemic), R&D should be supported to explore the potential of AI and GAI tools to detect misinformation from research articles and educational content.

- There should be a special emphasis on R&D to improve the reliability and competence of AI and GAI in medical education, scholarly medical writing, clinical diagnosis, and reporting.
- To ensure that extreme care is taken for medical education and research regarding the usage of AI and GAI tools, special ethical and legal guidelines should be issued, and strict compliance should be enforced.
- Data privacy and other ethical considerations related to the usage of AI for education, training, and healthcare should be addressed properly by the ethical and legal framework, violations should be seriously prosecuted, and offenders should be judiciously punished.
- The threat of AI replacing teachers, trainers, researchers and workers in some creative jobs is already there, so a proper work environment to ensure the coexistence of humans and AI tools should be ensured.
- It is to be noted that, while most of the clusters have grown substantially and a lot of new publications are added to the literature, most of the implications retrieved from analysis at phase 1 are found to be still relevant due to their immense gravity and multi-faceted complexity. The only things that have changed in this regard are the creation of new benchmarks for reliability, accuracy of performance and other aspects. Relevance of the above-mentioned implications may diminish only if there is advancement through a series of breakthroughs from synergetic coevolution of different disciplines of science & technology embracing socio-economic, environmental and ethical pillars of sustainable development. However, new themes that came into existence during phase 2 could reveal so many other relevant implications, but exploring those is beyond the scope of this work.
- With the discussion of implications, the sixth objective is also achieved. Thus, all the objectives are achieved with the proposed framework thereby demonstrating its effectiveness.

## CONCLUSION

As specified in the introduction, this work intended to develop a framework to determine the key fields of research, key research clusters, emerging research themes, and key research

contributions within grown and emerging clusters for themes that exhibit 'gold rush' kind of attention and growth characteristics as in the case of 'ChatGPT'. Various implications including scientific and technological implications, social implications including those related to the economy and job market, ethical implications including the ones related to the risk of life (possibly from medical applications), etc., that are useful to respective key stakeholders are also provided.

## Methodical contributions of this work

The framework built as a diligent combination of (i) science mapping using FoRs (provided by the Dimensions database) (ii) a fast community detection algorithm and (iii) the Flow Vergence model (powered for dynamic analysis), also enables analysts to track the development of early identified grown and emerging clusters. As all the objectives are successfully achieved using the framework, the framework is effective for analysis of the body of literature related to new but fast-growing (kind of gold rush) research topics like ChatGPT in its initial stage of development and tracking of the same after a suitable duration enabling the review of some of the insights gathered during the early analysis stage.

Also, in this work, science map creation from the Dimensions database is introduced, which is the first of this kind of attempt according to the best of our knowledge. Citation network creation using the code developed during this research (that uses unique publication IDs. provided by the Dimensions database) overcomes an existing limitation of VOSviewer. This also facilitated tracking of progress of grown and emerging clusters identified during first stage of analysis at the second stage. Thus, our contributed code supports dynamic analysis way better than VOSviewer.

## Limitations and directions for further research

In this framework, during first phase, determination of themes of grown and emerging clusters were done using mapping publications to FoRs. While this ensures the determination of these clusters' coarse or broad themes at very early stage, exploring ways to identify broad themes once early-detected emerging clusters grow sufficiently is a challenge. This is because, as network grows and cluster mergers happen, multi-themed clusters will be formed (as observed in second phase of our analysis). Also, determination of fine-level themes is required. Exploration of NLP techniques for addressing both of these can be really promising. Though Dimensions is one of the most available comprehensive databases, executing this framework on other databases might reveal some other important implications. The package developed for processing Dimensions data can be expanded to process data from other databases too. These are some of the enticing possibilities for further endeavours.

## ACKNOWLEDGEMENT

We want to express our immense gratitude to our beloved Chancellor, Mata Amritanandamayi Devi (AMMA), for providing the motivation and inspiration for this research work.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AVAILABILITY OF DATA AND MATERIALS

The data and materials are available with the authors and will be shared upon reasonable request communicated to the corresponding author via email.

## REFERENCES

- Haleem A, Javaid M, Singh RP. An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *Benchcounc Trans Benchmarks Stand Eval.* 2022;2(4):100089. doi: 10.1016/j.tbench.2023.100089.
- Daylight EG. Towards a historical notion of 'Turing-the father of Computer Science'. *Hist Philos Logic.* 2015;36(3):205-28. doi: 10.1080/01445340.2015.1082050.
- Roy K, Debdas S, Kundu S, Chouhan S, Mohanty S, Biswas B. Application of natural language processing in healthcare. *Comp Intell Healthc Inform.* 2021:393-407.
- Muscettola N, Nayak PP, Pell B, Williams BC. Remote agent: to boldly go where no AI system has gone before. *Artif Intell.* 1998;103(1-2):5-47. doi: 10.1016/S0004-3702(98)00068-X.
- Wen TH, Young S. Recurrent neural network language generation for spoken dialogue systems. *Comput Speech Lang.* 2020;63:101017. doi: 10.1016/j.csl.2019.06.008.
- Le QV. A tutorial on deep learning part 2: autoencoders, convolutional neural networks and recurrent neural networks. *Google Brain.* 2015;20:1-20.
- Amini N, Zhu Q. Fault detection and diagnosis with a novel source-aware autoencoder and deep residual neural network. *Neurocomputing.* 2022;488:618-33. doi: 10.1016/j.neucom.2021.11.067.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN. Attention is all you need. *Adv Neural Inf Process Syst.* 2017;30(1):261-72.
- Mars M. From word embeddings to pre-trained language models: A state-of-the-art walkthrough. *Appl Sci.* 2022;12(17):8805. doi: 10.3390/app12178805.
- Rahman MA, Saleh T, Jahan MP, McGarry C, Chaudhari A, Huang R, et al. Review of intelligence for additive and subtractive manufacturing: current status and future prospects. *Micromachines.* 2023;14(3):508. doi: 10.3390/mi14030508, PMID 36984915.
- Kolides A, Nawaz A, Rathor A, Beeman D, Hashmi M, Fatima S, et al. Artificial intelligence foundation and pre-trained models: fundamentals, applications, opportunities, and social impacts. *Simul Modell Pract Theor.* 2023;126:102754. doi: 10.1016/j.simpat.2023.102754.
- Cooper G. Examining science education in ChatGPT: an exploratory study of generative artificial intelligence. *J Sci Educ Technol.* 2023;32(3):444-52. doi: 10.1007/s10956-023-10039-y.
- Nalimov VV, Mulchenko ZM. Measurement of science. Study of the development of science as an information process.
- Lathabai HH, Prabhakaran T, Changat M. Contextual productivity assessment of authors and journals: a network scientometric approach. *Scientometrics.* 2017;110(2):711-37. doi: 10.1007/s11192-016-2202-0.
- Prabhakaran T, Lathabai HH, Changat M. Detection of paradigm shifts and emerging fields using scientific network: A case study of Information Technology for Engineering. *Technol Forecasting Soc Change.* 2015;91:124-45. doi: 10.1016/j.techfore.2014.02.003.
- Singh VK, Singh P, Karmakar M, Leta J, Mayr P. Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics.* *J Coverage of.* 202;126:5113-42.
- Borgatti SP. Centrality and network flow. *Soc Netw.* 2005;27(1):55-71. doi: 10.1016/j.socnet.2004.11.008.
- Waltman L, Van Eck NJ, Noyons EC. A unified approach to mapping and clustering of bibliometric networks. *J Inf.* 2010;4(4):629-35. doi: 10.1016/j.joi.2010.07.002.
- Seidman SB. Network structure and minimum degree. *Soc Netw.* 1983;5(3):269-87. doi: 10.1016/0378-8733(83)90028-X.
- Zaveršnik M, Batagelj V. Islands. Slides from sunbelt XXIV. Portoroz, Slovenia. 2004;12:16.
- Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theor Exp.* 2008;2008(10):10008. doi: 10.1088/1742-5468/2008/10/P10008.
- Prabhakaran T, Lathabai HH, George S, Changat M. Towards prediction of paradigm shifts from scientific literature. *Scientometrics.* 2018;117(3):1611-44. doi: 10.1007/s11192-018-2931-3.
- Prabhakaran T, Lathabai HH, George S. Competing, complementary and co-existing paradigms in techno-scientific literature: a case study of nanotechnology for engineering. *Scientometrics.* 2019;118(3):941-77. doi: 10.1007/s11192-019-03013-2.
- Lathabai HH, Prabhakaran T, Changat M. Centrality and flow vergence gradient based path analysis of scientific literature: A case study of biotechnology for engineering. *Phys A.* 2015;429:157-68. doi: 10.1016/j.physa.2015.01.085.
- Lathabai HH, George S, Prabhakaran T, Changat M. An integrated approach to path analysis for weighted citation networks. *Scientometrics.* 2018;117(3):1871-904. doi: 10.1007/s11192-018-2917-1.
- Aljanabi M, Ghazi M, Ali AH, Abed SA. ChatGPT: open possibilities. *Iraqi J Comput Sci Math.* 2023;4(1):62-4.
- Taecharungroj V. "What can ChatGPT do?" Analyzing early reactions to the innovative AI chatbot on Twitter. *Big Data Cogn Comput.* 2023;7(1):35. doi: 10.3390/bdccc7010035.
- Wang FY, Miao Q, Li X, Wang X, Lin Y. What does ChatGPT say: the DAO from algorithmic intelligence to linguistic intelligence. *IEEE CAA J Autom Sin.* 2023;10(3):575-9. doi: 10.1109/JAS.2023.123486.
- Mann DL. Artificial intelligence discusses the role of artificial intelligence in translational medicine: a JACC: basic to translational science interview with ChatGPT. *JACC Basic Transl Sci.* 2023;8(2):221-3. doi: 10.1016/j.jaccbts.2023.01.001, PMID 36908674.
- Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. *Lancet Digit Health.* 2023;5(4):179-81. doi: 10.1016/S2589-7500(23)00048-1, PMID 36894409.
- Yeo YH, Samaan JS, Ng WH, Ting PS, Trivedi H, Vipani A, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol.* 2023;29(3):721-32. doi: 10.3350/cmh.2023.0089, PMID 36946005.
- Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLoS Digit Health.* 2023;2(2):0000198. doi: 10.1371/journal.pdig.0000198, PMID 36812645.
- Liebrezn M, Schleifer R, Buadze A, Bhugra D, Smith A. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. *Lancet Digit Health.* 2023;5(3):105-6. doi: 10.1016/S2589-7500(23)00019-5, PMID 36754725.
- Patel SB, Lam K, Liebrezn M. ChatGPT: friend or foe? *Lancet Digit Health.* 2023;5(3):102. doi: 10.1016/S2589-7500(23)00023-7, PMID 36754723.
- Stokel-Walker C. AI bot ChatGPT writes smart essays-should academics worry? *Nature.* 2022;603(7902):563-. doi: 10.1038/d41586-022-00620-7.
- Gao CA, Howard FM, Markov NS, Dyer EC, Ramesh S, Luo Y, et al. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *bioRxiv.* 2022 Dec 27: 2022-12.
- Aydın Ö, Karaarslan E. OpenAI ChatGPT generated literature review: digital twin in healthcare. Aydın, Ö. Karaarslan, E; 2022. Open AI. ChatGPT generated literature review: digital twin in healthcare. In: Aydın Ö, editor. *Emerging Computer Technologies.* Vol. 2;2022 Dec 21.
- Biswas SS. Role of Chat GPT in public health. *Ann Biomed Eng.* 2023;51(5):868-9. doi: 10.1007/s10439-023-03172-7, PMID 36920578.
- Biswas SS. Potential use of Chat GPT in global warming. *Ann Biomed Eng.* 2023;51(6):1126-7. doi: 10.1007/s10439-023-03171-8, PMID 36856927.
- Will ChatGPT transform healthcare? *Nat Med.* 2023;29(3):505-6. doi: 10.1038/s41591-023-02289-5, PMID 36918736.
- Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, Chartash D. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ.* 2023;9(1):e45312.
- Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. *J Educ Eval Health Prof.* 2023;20(1):1. doi: 10.3352/jeehp.2023.20.1, PMID 36627845.
- Cahan P, Treutlein B. A conversation with ChatGPT on the role of computational systems biology in stem cell research. *Stem Cell Rep.* 2023;18(1):1-2. doi: 10.1016/j.stemcr.2022.12.009, PMID 36630899.
- Thorp HH. ChatGPT is fun, but not an author. *Science.* 2023;379(6630):313-. doi: 10.1126/science.adg7879, PMID 36701446.
- Zhavoronkov A. Rapamycin in the context of Pascal's Wager: generative pretrained transformer perspective. *Oncoscience.* 2022;9:82-4. doi: 10.18632/oncoscience.571, PMID 36589923.
- Else H. Abstracts written by ChatGPT fool scientists. *Nature.* 2023;613(7944):423. doi: 10.1038/d41586-023-00056-7, PMID 36635510.
- Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature.* 2023;613(7945):620-1. doi: 10.1038/d41586-023-00107-z, PMID 36653617.
- The AI Writing on the wall. *Nat Mach Intell.* 2023;5(1):1-. doi: 10.1038/s42256-023-00613-9.

49. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. *Nature*. 2023;614(7947):214-6. doi: 10.1038/d41586-023-00340-6, PMID 36747115.
50. Krügel S, Ostermaier A, Uhl M. ChatGPT's inconsistent moral advice influences users' judgment. *Sci Rep*. 2023;13(1):4569. doi: 10.1038/s41598-023-31341-0, PMID 37024502.
51. Hoorn JF, Chen JJ. Epistemic considerations when AI answers questions for us. *arXiv preprint arXiv:2304.14352*. 2023 Apr 23.
52. Haase J, Hanel PH. Artificial muses: generative artificial intelligence chatbots have risen to human-level creativity. *J Creativity*. 2023;33(3):100066. doi: 10.1016/j.yjoc.2023.100066.
53. Frosio G. The artificial creatives: the rise of combinatorial creativity from dall-e to gpt-3. *Handbook of artificial intelligence at work: interconnections and policy implications*.
54. Farhat F, Sohail SS, Madsen DØ. How trustworthy is ChatGPT? The case of bibliometric analyses. *Cogent Eng*. 2023;10(1):2222988. doi: 10.1080/23311916.2023.2222988.
55. Kirtania DK. ChatGPT as a tool for bibliometrics Analysis: interview with ChatGPT. Available at SSRN 4391794. *SSRN Journal*. 2023 Mar 17. doi: 10.2139/ssrn.4391794.
56. Kirtania DK. Network Visualization of ChatGPT Research: a study based on term and keyword co-occurrence network analysis. *arXiv preprint arXiv:2304.01948*. 2023.
57. Hoes E, Altay S, Bermeo J. Using ChatGPT to fight misinformation: ChatGPT nails 72% of 12,000 verified claims. *PsyArXiv*. April. Vol. 3; 2023 .
58. Reiss MV. Testing the reliability of chatgpt for text annotation and classification: A cautionary remark. *arXiv preprint arXiv:2304.11085*. 2023.
59. Li L, Fan L, Atreja S, Hemphill L. Hot. ChatGPT: The promise of ChatGPT in detecting and discriminating hateful, offensive, and toxic comments on social media. *ACM transactions on the web*. 2024;18(2):1-36.
60. Huang F, Kwak H, An J. Is ChatGPT better than human annotators? potential and limitations of ChatGPT in explaining implicit hate speech. In *Companion Proc ACM web conference*. Vol. 2023; 2023. p. 294-7. doi: 10.1145/3543873.3587368.
61. Gupta R, Herzog I, Weisberger J, Chao J, Chaiyasate K, Lee ES. Utilization of ChatGPT for plastic surgery research: friend or foe? *J Plast Reconstr Aesthet Surg*. 2023;80:145-7. doi: 10.1016/j.bjps.2023.03.004, PMID 37023599.
62. Kleebayoon A, Wiwanitkit V. Utilization of ChatGPT for plastic surgery research: comment. *J Plast Reconstr Aesthet Surg*. 2023;82:275. doi: 10.1016/j.bjps.2023.04.084, PMID 37209600.
63. McGee RW. How would American history be different if FDR had been assassinated in 1933: A ChatGPT essay. Available at SSRN 4413419. *SSRN Journal*. 2023. doi: 10.2139/ssrn.4413419.
64. McGee RW. Is ESG a bad idea? The ChatGPT response. *SSRN Journal*. Apr 8, 2023; doi: 10.2139/ssrn.4413431.
65. McGee RW. How to Determine Your Fair Share of Taxes: Ask Chatgpt. Available at SSRN 4413435. 2023.
66. Khademi A. Can ChatGPT and bard generate aligned assessment items? A reliability analysis against human performance. *arXiv preprint arXiv:2304.05372*. 2023.
67. Watkins R. Guidance for researchers and peer-reviewers on the ethical use of Large Language Models (LLMs) in scientific research workflows. *AI Ethics*. 2023;1-6. doi: 10.1007/s43681-023-00294-5.

**Cite this article:** Lathabai HH, Prabhakaran T, Raman R. ChatGPT Research: Insights from Early Studies Using Network Scientometric Approach. *J Scientometric Res*. 2024;13(3):688-705.