Visual Storytelling: Text-to-Video Conversion from Bibliometric Perspectives

Gaganpreet Kaur[1], Amandeep Kaur[1], Meenu Khurana[2]

Abstract: Building upon previous research on Text2Image, Text2Presentation, and Text2Video, this study adds value by critically evaluating a wide range of current methodologies, including different generative models, critical analysis of various approaches, highlighting the shortcomings, and suggesting and directing future research. Text-to-video generation is a cutting-edge and quickly developing field of artificial intelligence that uses computer vision and natural language processing to generate video material from textual descriptions.A popular and rigorous method for searching and assessing large volumes of scientific material is bibliometric analysis. As a result, we aim to provide a comprehensive overview of the bibliometric approach, emphasizing its various methodologies, along with detailed instructions that can be trusted to ensure that bibliometric analysis is carried out rigorously and confidently. To this end, we elucidate when and where to apply bibliometric analysis in relation to related techniques such as systematic literature reviews and meta-analyses. This work presents the first bibliometric analysis of text-to-video production, filling a significant gap in the literature by providing a comprehensive overview of major themes, prominent researchers, and developing areas within this rapidly evolving field. Taken as a whole, this research should be a useful tool for comprehending the many methods and approaches that can be used to conduct bibliometric analysis research.

Keywords: Text2Video, Education, Research, Bibliometric, Precision, adaptive storytelling.

1. Introduction

In the current digital era, text-based communication of ideas and concepts is not always sufficient. Since visual learners, using visual aids can improve comprehension and retention. As a result, the capacity to communicate information visually has grown in significance. This article will examine the evolution of transition from text-based explanations to video representations [1][2].

The use of visual aids has grown much more common since the internet and other digital media were invented. Visual aids come in a variety of forms today, including posters, slides, images, infographics, animations and videos. These graphic aids assist in simplifying and illuminating difficult ideas and concepts [3].

In the realm of artificial intelligence, text to video generation plays an important role in marketing, teaching, and entertainment and text to video generation has recently been a hot topic. It entails the automatic conversion of text into video information, which can be utilised for a variety of things. A wide range of applications, including video production, education, marketing, and entertainment, have the potential to be transformed by this technology. The bibliometric review of the literature on text to video

generation and its application to enhancing online visibility and engagement will be covered in this article [4].

Over the course of time, scholars have persistently enhanced and perfected the first text-to-video generating systems by the utilisation of many methodologies, including deep learning, computer graphics, and motion analysis[7]. In contemporary times, there exists a diverse array of tools and platforms that facilitate the creation of movies through textual inputs. These tools span from rudimentary text-to-image converters to sophisticated systems driven by artificial intelligence [5][6][8]. This analysis will show potential directions for this promising technology's development and offer insightful information on the state of research.

1.1. Objective

This systematic and bibliometric evaluation aims to offer a thorough and all-encompassing analysis of the current advancements in text-to-video generation. More precisely, this review aims to tackle the subsequent research inquiries: This research focuses to examine documents published over the last approx. 24 years for detailed bibliometric survey for growing trends in text to video research and its applications in many sectors.

2. Research Methods and Information Collection

Three well-known bibliographic databases—Scopus, Lens, and Dimension—were first chosen in order to

[1,1]* *Chitkara Institute of Engineering and Technology Chitkara University*
* [2]*Chitkara School of Engineering and Technology Chitkara University*
* *Corresponding Author Email: gaganpreet.kaur@chitkara.edu.in*

do a thorough bibliometric analysis. Comparing the publication trends in different databases allowed us to determine which one would be best for a more in-depth examination [19].

The given Table [1] presents the yearly count of publications indexed in three bibliographic databases, namely Scopus, Lens, and Dimension, between the years 2000 and 2025. A consistent and thorough coverage of peer-reviewed literature in Fig.1 is demonstrated by Scopus, which is represented by the blue line and indicates a steady growth in publications that peaks about 2022–2023. Orange-colored Lens, which peaks around 2022, has an identical increasing tendency but fewer articles overall. A notable peak in publications around 2021–2022 is seen in the dimension, represented by the green line, which is followed by a steep decrease, indicating possible discrepancies. For bibliometric analysis, Scopus was chosen because of its broad coverage, dependable and high-quality data, sophisticated citation analysis capabilities, and steady publication trends. Lens and Dimension, on the other hand, while helpful, do not provide the same degree of steadiness and comprehensiveness, making Scopus.

Table 1: Yearly Count Publications

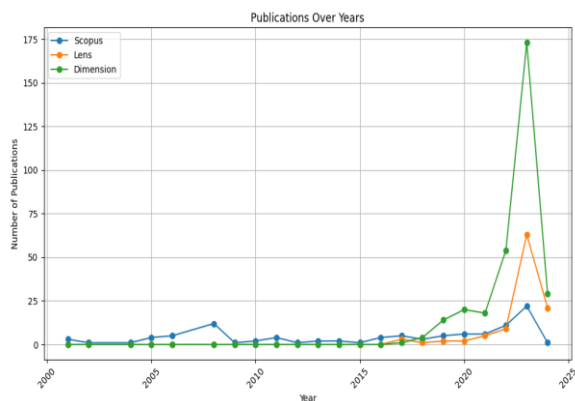| Year	Scopus	Lens	Dimension
2001	3.0	0.0	0.0
2002	1.0	0.0	0.0
2004	1.0	0.0	0.0
2005	4.0	0.0	0.0
2006	5.0	0.0	0.0
2008	12.0	0.0	0.0
2009	1.0	0.0	0.0
2010	2.0	0.0	0.0
2011	4.0	0.0	0.0
2012	1.0	0.0	0.0
2013	2.0	0.0	0.0
2014	2.0	0.0	0.0
2015	1.0	0.0	0.0
2016	4.0	0.0	0.0
2017	5.0	3.0	1.0
2018	3.0	1.0	4.0
2019	5.0	2.0	14.0
2020	6.0	2.0	20.0
2021	6.0	5.0	18.0
2022	11.0	9.0	54.0
2023	22.0	63.0	173.0
2024	1.0	21.0	29.0

Fig.1 Publication over Years

Since the Scopus database is the largest peer-reviewed books, journals, and conference proceedings, it served as the source of data. With 157 published publications, the survey covers a total of 23.5 years, or from 2000 to May 2024. The comprehensive process is shown graphically in Fig.1 The research study includes the following: top-cited documents, titles, keywords, abstract terms, sources, affiliations, count of publications, types of published articles, and analysis of top authors. Word clouds of the top 100 often used terms from the title, keywords, and abstract have also been plotted according to word frequency for improved comprehension through visualisation. A summary of challenges and applications of text to video generation is also provided[9][10].

3. Methodological Framework

This study uses a systematic bibliometric analysis on text-to-video generation, utilizing advanced data analysis tools and a comprehensive data collection process. It examines research trends, collaboration networks, and citation patterns using specialized software like Bibliometric and Vosviewer. This structured approach provides insights into the current landscape and future directions of text-to-video generation research.

3.1. Data analysis tools

3.1.1. Bibliometric

The R programming language and a logical bibliometric approach are used by the Bibliometric application to carry out a variety of analysis, such as co-citation, coupling, and scientific cooperation. Bibliographic information can be imported from databases such as PubMed, and Scopus. Bibliometric analysis estimates productivity and impact while assisting researchers in understanding publication output, citation trends, collaboration networks, and research effects[11].

3.1.2. Vosviewer

Leiden University researchers created the user-friendly VOSviewer program, which is used to analyze bibliometric networks. Users can use it to map scientific literature based on relationships between co-citations. It can alter maps and analyze different kinds of data[12].

3.2. Data Collection

An electronic database is a sorted collection of data that can be kept and accessed. Multimedia data can be managed and stored in databases along with text, numbers, images, and other kinds of data. They are widely used in businesses, governments, and organizations to evaluate enormous volumes of data. A significant amount of scholarly papers' references and abstracts are available in a database called Scopus. It covers a broad spectrum of academic disciplines, including technology, arts, social sciences, and medical sciences. It is owned and maintained by Elsevier, one of

the largest academic publishers. With about 25 million patents, it contains about 75 million records. Scholars can monitor, evaluate, and analyze research trends with the aid of several tools offered by Scopus. Users can identify important research trends and new fields of investigation with the use of these approaches, which include author profiling, co-citation networks, and citation analysis. Choosing a search strategy and gathering the necessary data are the two phases that comprise the methodological part of the study.This work uses a bibliographic evaluation that is based on the Scopus database (www.scopus.com) "Text-to-Video Generation" OR "Video Generation from Text" OR "Text-Based Video Synthesis" OR "Automatic Video Generation from Text" OR "Natural Language Video Synthesis" OR "Text-Driven Video Creation" OR "Video Generation via Natural Language Processing" OR "Text-to-Visual Content Generation" OR "Video Generation Using Textual Input" OR "Language-Based Video Production" The papers in the dataset were found by utilizing the search phrases from the database in the search title. When all publications are considered, 157 documents are found, and when only articles are looked at, 85 matches are found.[12].
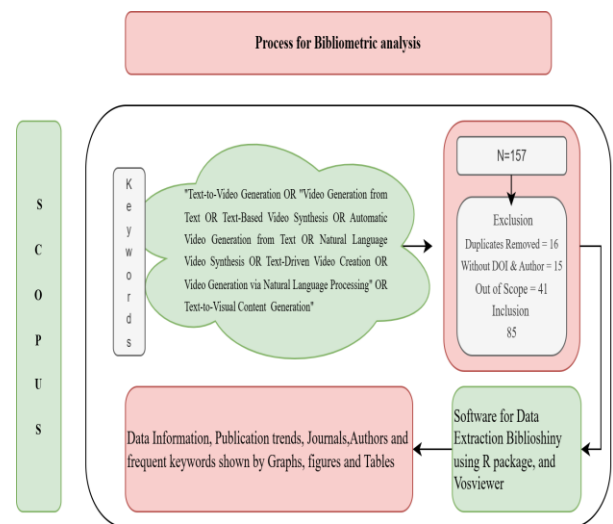

Fig.2 Workflow of Bibliometric Analysis

4. Research Questions

Research Questions	Objective	Motivation
RQ1 What are the applications of Text to Video generation?	To convert written content into visually appealing and captivating video content for a range of applications in different fields and industries.	Transform written content into eye catch ing video content to impro engagement and Communication.
RQ2 Describe the main approaches used in Text to Video generation?	To investigate and clarify the main strategies and tactics used in the transformation of written content into video.	Improve knowledge of the procedu entailed in the conversion of text to video.
RQ3: What are the main bibliometric statistics and their effect in the research landscape?	Examine research methods for Text to Video Generation, list the top scientists in each nation, and look up commonly used terms.	Learn about research methodologies, identify the top nations in the subject, and pinpoint common terms to direct future studies and conversations.
RQ4 What are the main challenges and scope for future research areas?	Determine the main issues and directions for further study in the field of text to video production.	The goal of the study is to discover areas for development in order to produce multimedia material that is more engaging by examining the efficacy and strengths of text-to- video production techniques.

4.1. Applications of Text-to-Video Generation

RQ1: What are the applications of Text- to-Video Generation?

- **Enhanced Communication and Engagement:** Text-to-video production improves audience engagement in marketing, corporate presentations, and educational contexts by turning textual content into dynamic visual display [2].

- **Transferring Knowledge Effectively:** Since visually appealing videos may be accessible from any location, text-to-video creation improves information transfer in training and education while saving time and effort especially for e-learning platforms[2].

  Creative Storytelling & Marketing: With the help of text-to-video generation, businesses can turn product descriptions, customer testimonials, and brand tales into interesting videos that successfully convey their message and raise brand awareness, sales, and customer engagement [3].

- **Production of Tailored Content:** With the use of text-to-video creation, con- tent creators may tailor their works to specific user profiles or audiences. By adding voiceovers, animations, and graphics to their content, users can become more engaged and enjoy it more [4].

- **Media and Entertainment Production:** With the advent of virtual worlds, animated storytelling, and special effects in movies, television shows, and online streaming platforms, text-to-video generation has completely transformed the entertainment sector [5].
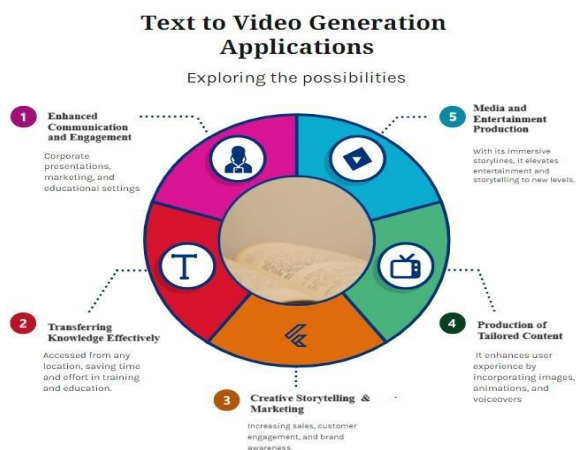

Fig. 3 Applications of Text to Video

4.2. Approaches used in Text-to-Video generation

RQ2: Describe main approaches used in text to video

generation?

- In the 2010 to 2014, researchers began to explore the concept of text-to-video creation, wherein animations may be generated from textual descriptions. The initial research efforts in this field were mostly centered around rule-based methodologies and Template matching wherein researchers would manually define the correspondence between textual information and visual components while template matching utilized pre- recorded video clips to match with text input [4].

- During the 2010s, there were notable developments in the field of text-to-video production. Researchers focused on investigating deep learning methodologies, including neural networks and generative adversarial networks (GANs). These techniques have facilitated the ability of computers to generate movies that are more realistic and intricate based on textual descriptions. As a result, new opportunities have emerged for the utilisation of these techniques in various applications, including video summarization and video captioning [5].

- Most recent developments have concentrated on using more complex deep learning methods—like transformers and multi-modal learning—to improve text-to-video production even more. To enhance video generation, multi-modal techniques incorporate several data kinds, including text, audio, and visual information. Transformer models are designed to capture long-range dependencies between video frames and text inputs by utilizing self-attention techniques. BERT and GPT are two examples of pre-trained language models that are used to improve text comprehension and direct the creation of videos [8].
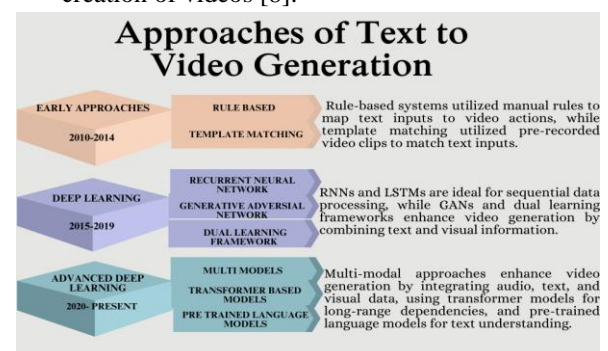

Fig.4 Approaches of Text to Video Generation

In general, the evolution of text-to-video creation may be observed as a transition from initial rule-based methodologies to more sophisticated data-driven methodologies, facilitated by developments in machine learning and natural language processing methodologies. The development of this technology has presented novel opportunities for the creation of superior video content based on textual descriptions. These advancements hold promise for various domains, including but not limited to video production, advertising, and education.

Researching this technology will have an impact on editing, security, communication, and other technological issues. It will allow devices to record live video and produce a trimmed version, which will improve digital experiences everywhere. Many techniques, including video models, deep video generation, GANs, conditional flow-based models, formal logic, and latent neural differential equations, are used in automatic video creation. Innovative techniques are being investigated to increase training effectiveness and address problems with video production [6]. In this study examines advanced AI methodologies in Text-to-Image and Text-to-Video generations, examining neural networks, data preprocessing methods, and field-use evaluation. It highlights challenges, limitations, and future research directions, highlighting their potential in digital marketing, content creation, and video production [7].

4.3. Exploring the major Bibliometric statistics and research priorities.

RQ3: What are the main bibliometric statistics and their effect in the research landscape?

Bibliometrics is a research method that counts articles, citations, and reviews to assess an institution's or organization's success. It helps identify topics for future research and provides writers with recommendations for more investigation. There are two basic strategies that focus on understanding the relationships between study elements: scientific mapping and performance analysis.

4.3.1. Dataset Information

Fig. 5 Dataset basic Information

The basic description of 84 documents that were taken from the Scopus database and further used in this manuscript is shown in Fig. 2. Primary data, document contents, authors, authors' collaborations, and document categories have all been used to categorize the information that was retrieved. There are 2724 references in all for the chosen topic, including 60 sources from books and journals. Each work has an average of 9.06 citations to other articles and an 8.83% annual growth rate. There are 313 authors total and 209 distinct author

keywords in the papers. 4.14 co-authors and 16.67 international co-authors collaborated on the data collection, while five people worked alone. The articles are divided into several groups, including review papers, conference papers, and book chapters.

4.3.2. Publication Per Year

To evaluate the evolution of text to video generation research, data spanning 24 years—from 2001 to 2024—has been gathered. The graph that goes with the study shows in Fig.6 how the field of video generation research is always changing. Although the precise number of publications for 2024 has not yet been determined (because the search is still ongoing), it is evident that the topic has garnered more attention lately and is still relatively new.
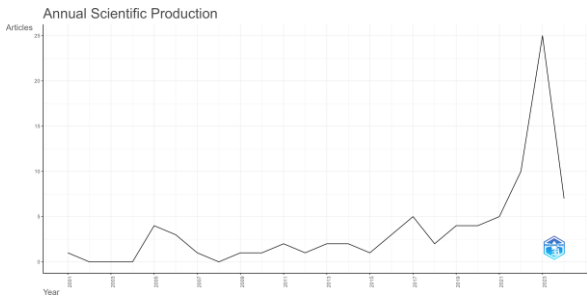

Fig.6 Annual Publication

4.3.3. Threefold Plot

The Three plot diagram visually represent the distribution of study participants across countries and the recentness of referred references. The graph shown in Figure 5 lists the ten title keywords commonly used by researchers in text to video research, such as "multimodal," "visual," "text to video," and "content." The top ten approaches for text to video include "Deep Learning," "video generation," "text to video generation," "GAN," and "content-based"[21][22].The study examines the ten main contributing countries, including China, India, the United States, Korea, Hong Kong, and Germany, and discusses the interaction between these three sectors. The choice of research methodologies is greatly influenced by the choice of title keywords, which is crucial for deciding the future course of research and developing effective decision support systems.
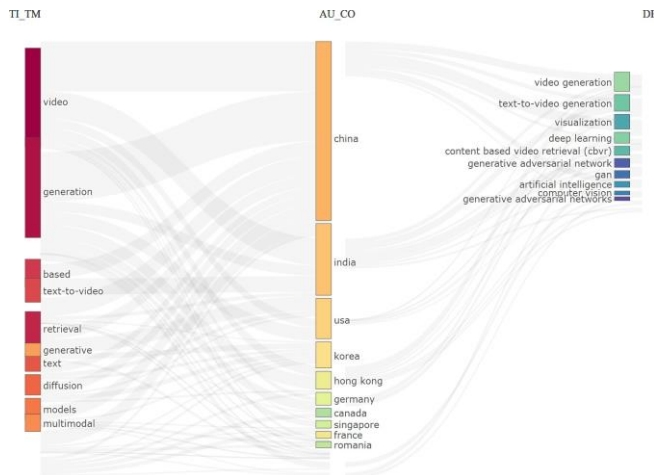
Fig. 7 Threefold Plot: Titles, Countries & Keywords

4.3.4. Most relevant Sources

The study highlights the importance of using influential and frequently referenced sources in a specific study field. It highlights the significance of "Lecture Notes in Computer Science" as the most significant source. Other important sources include "Proceedings of the IEEE International Conference on Computer Vision", "Advances in Neural Information Processing Systems", "Computer Vision and

Fig. 8 Most Relevant Sources

Pattern Recognition", and "IEEE Transactions on Multimedia". The results emphasize the importance of using a variety of sources to fully understand a study issue shown in Fig.8.

4.3.5. Bardford Law

The distribution of scientific material among journals is explained by the bibliometric theory known as the Law of Bradford. According to this, the quantity of papers published in a journal is inversely correlated with its ranking, meaning that higher ranked journals publish a greater number of articles. Samuel C. Bradford noticed this pattern in the distribution of papers in geological publications. The source generated by "Lecture Notes in Computer Science" with the greatest number of citations is displayed in Fig.9. It is ranked#1 and belongs to Zone 1, with a frequency of 12 and a cumulative frequency of 12. Other publications include "Proceedings of the IEEE International Conference on Computer Vision," "ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS," "IEEE Access," and the "Proceedings of the AAAI Conference on Artificial Intelligence."[17].

SO	Rank	Freq	cumFreq	Zone
LECTURE NOTES IN COMPUTER SCIE	1	12	12	Zone 1
PROCEEDINGS OF THE IEEE INTERN/	2	4	16	Zone 1
ADVANCES IN NEURAL INFORMATIC	3	3	19	Zone 1
PROCEEDINGS OF THE AAAI CONFEF	4	3	22	Zone 1
PROCEEDINGS OF THE IEEE COMPU'	5	3	25	Zone 1
ACM INTERNATIONAL CONFERENCE	6	2	27	Zone 1
IEEE TRANSACTIONS ON MULTIMED	7	2	29	Zone 1
PROCEEDINGS - IEEE INTERNATION/	8	2	31	Zone 2
PROCEEDINGS OF THE ANNUAL MEI	9	2	33	Zone 2
2006 IEEE INTERNATIONAL CONFER	10	1	34	Zone 2

Fig. 9 Law of Bardford Cited Sources

4.3.6. Yearly Production of sources

Source production over time in bibliometric analysis refers to the number of publications produced by a single author or institution over a specific period. This information is derived from bibliographic data from databases like Dimension, Web of Science, and Scopus. The Fig.10 shows a constant upward trend in article creation from 2017 to 2024, indicating significant research activity influenced by factors like funding increases, new study topics, and technology breakthroughs.
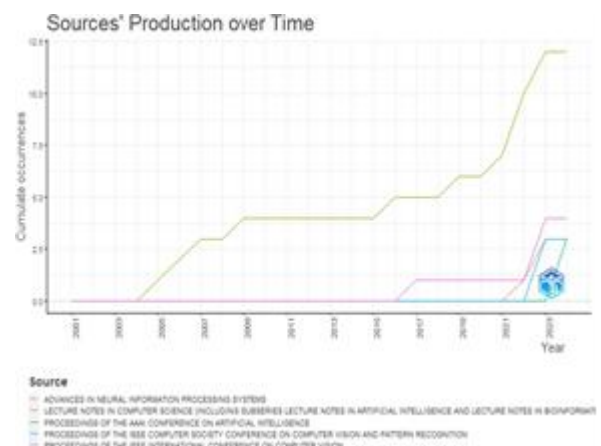

Fig.10 Yearly Source Production

4.3.7. The Top Authors

In bibliometrics, the phrase "most significant author" designates a person who has had a major impact on a

certain topic, ascertained mostly by citation analysis and bibliographic data. This analysis considers metrics like publications, citation counts, and h-index to determine the most important authors in a subject. In a graph, "Clavel, Choé," "Thape, Sudep," "Vilhjalmsson, Hans," and "Wang, Xintao" have the most publications, indicating their significant influence in the Text to Video Production sector.
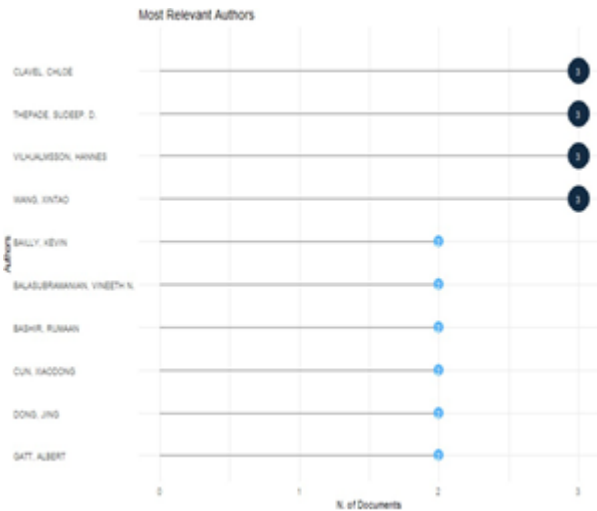

Fig. 11 Most Relevant Authors

4.3.8. The Top Affiliations

The Most Significant Affiliation (MSA) of research output is determined by bibliometric tools such as the Fractional Count approach and the Affiliation Index, which examine author affiliations in publications. Korea University and Peking University have the most affiliation in a specific field, while South China University of Technology has the least association, publishing only four articles in the same subject area.
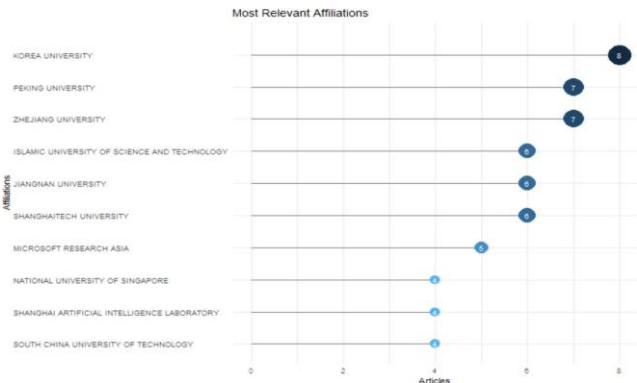

Fig. 12 Most Relevant Affiliations

4.3.9. Most Cited Countries

5. The countries that have made the biggest contributions to the corpus of knowledge on a certain subject are those included under the heading "Most Significant Cited Countries" (MSCC). The MSCC designation is based on how frequently researchers from that country are

cited in scholarly works in that field of study. The citation breakdown by country is provided in Fig.13. It shows that"Germany" emerged as the top country with the most citations, with contributions to 275 published works. With 107 citations, "China" ranked second, ahead of "Korea," "USA," "UK," "France," and "Australia."
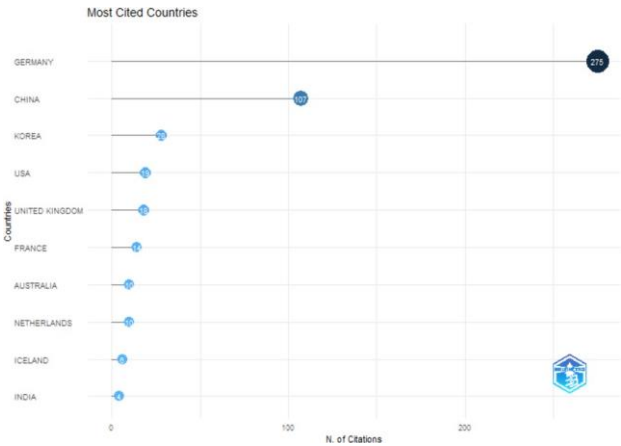

Fig. 13 Most Cited Countries

5.1.1. Most Frequently used Terms

It is a text analysis technique that removes punctuation and stopword to identify frequently occurring words. It evaluates textual data points, such as full-text articles or abstracts, to identify popular themes, key words, or research areas. In this study, "Video generation" and "Multi-modal" were found to have the highest frequency of 26 and 16.
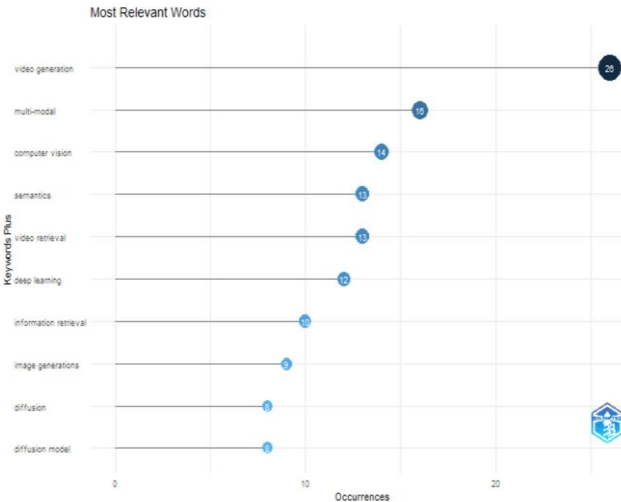

Fig. 14 Frequently used Words

5.1.2. WordCloud

In bibliometrics, word clouds are used to explore main topics, subjects, and study areas within a discipline. Word clouds are a visual depiction of the most commonly used terms in a dataset or corpus. By choosing 100 keywords based on the author and title

keywords, two distinct word clouds were produced [13].

Fig. 15 Wordcloud basis on author Keywords

Fig. 16 Wordcloud basis on Title

5.1.3. TreeMap

A treemap is a visual tool [15]that displays hierarchical data in rectangle format, with each rectangle representing a different value or quantity. The graphic displays multiple keywords, each represented by a different colored rectangular image. Examples include "video generation" and "multi-modal" in Fig. 17.

Fig. 17 Title based Tree Map

5.1.4. Co-occurence analysis of keywords

Analyzing the co-occurrence of keywords can reveal underlying themes and sub- jects, and finding relationships between phrases can refine search terms and improve keyword-based searches accuracy in Fig.16. Cluster analysis is a useful technique for organising related objects or entities based on shared characteristics. In network visualisation maps, cluster analysis is used to identify and display node clusters

that are closely associated to one another as distinct clusters. The study found that just 26 out of the 716 total keywords met the cutoff limit of at least 5 occurrences. By further classifying these 26 keywords into 4 groups, the connections and correlations between them were emphasised. Fig.18 shows four main clusters as "Red," "Green," "Blue," and "Yellow" bubbles. There are eight in the red colour bubble, such as "Multi-modal," "Computer Vision," and "Computational Linguistics." The green colour bubble has seven components, such as "Artificial intelligence," "Deep learning," "Video generation," etc. The blue colour bubble contains six objects, such as "Image retrieval," "Content based video retrieval," and so on. There are five items in the yellow colour bubble, such as "text to video generation" and "visualization"[14].

Fig. 18 Analysis of keywords

6. Challenges and Scope

RQ4: What are the main challenges and scope for future research

areas?

Text-to-video creation faces several challenges, including the lack of high-quality paired text-video datasets, difficulties in understanding complex written descriptions, and the need for improved contextual comprehension. To overcome these issues, researchers should create large-scale, diverse datasets, use natural language processing algorithms, and improve temporal coherence among video frames. Increasing realism can be achieved through advanced generative models and multi-modal data. The computational complexity of deep learning model training necessitates the development of more effective algorithms and improved model structures. The lack of established evaluation standards for video quality necessitates the creation of comprehensive evaluation frameworks considering multiple dimensions. Future research should

focus on cross-lingual and cross-cultural video generation, adaptive and personalized video generation, real-time video generation, ethical and societal consequences, and interactive systems. A multidisciplinary strategy involving Deep-Learning, Natural Language Processing, Computer Vision, and human-computer- interaction is needed. Cooperation between industry, academics, and policymakers is essential for text-to-video generation to reach its full potential [18].

7. Conclusion and future Work

In this bibliometric research article, we examined publishing patterns, prominent authors, top institutions, and important research issues in our chosen field using the Scopus database and Biblioshiny software. Our analysis emphasised important contributions and institutions and showed a notable increase in publications, suggesting a growing research community. The field is dynamic, and new patterns were highlighted by the theme analysis. To improve research depth and impact, future work should concentrate on interdisciplinary collaborations, emerging technologies, international partnerships, and advanced data analytics. To increase the accuracy and usefulness of bibliometric studies, however, issues including data accessibility and quality, a lack of standardised approaches, disciplinary obstacles, and the requirement for technology adaption must be resolved. All things considered, this study provides insightful information, but in order to advance the area further, these obstacles must be overcome and other avenues must be investigated.

References

[1] K. Paramita and M. L. Khodra, "Tailored summary for automatic poster generator," in *2016 International Conference on Advanced Informatics: Concepts, Theory and Application (ICAICTA)*, 2016, pp. 1–6.

[2] J. Oppenlaender, "The creativity of text-to-image generation," in *Proc. 25th Int. Acad. Mindtrek Conf.*, 2022, pp. 192–202.

[3] C. Xu, R. Wang, S. Lin, X. Luo, B. Zhao, L. Shao, and M. Hu, "Lecture2note: Automatic generation of lecture notes from slide-based educational videos," in *2019 IEEE Int. Conf. Multimedia Expo (ICME)*, 2019, pp. 898–903.

[4] Y. Li, M. Min, D. Shen, D. Carlson, and L. Carin, "Video generation from text," in *Proc. AAAI Conf. Artif. Intell.*, vol. 32, 2018.

[5] Y. Yu, Z. Tu, L. Lu, X. Chen, H. Zhan, and Z. Sun, "Text2Video: Automatic video generation based on

text scripts," in *Proc. 29th ACM Int. Conf. Multimedia*, 2021, pp. 2753–2755.

[6] R. K. Chauhan, G. Kalia, and A. Mishra, "Video editing: Technical vs. creative tool," *ECS Trans.*, vol. 107, no. 1, p. 10045, 2022.

[7] G. Kaur, A. Kaur, M. Khurana, et al., "A stem to stern sentiment analysis emotion detection," in *2022 10th Int. Conf. Reliability, Infocom Technol. Optimization (Trends Future Directions) (ICRITO)*, 2022, pp. 1–5.

[8] G. Kaur, A. Kaur, and M. Khurana, "A survey of computational techniques for automated video creation and their evaluation," in *2024 11th Int. Conf. Reliability, Infocom Technol. Optimization (Trends Future Directions) (ICRITO)*, 2024, pp. 1–6.

[9] A. S. Rusydiana, "Bibliometric analysis of journals, authors, and topics related to covid-19 and islamic finance listed in the dimensions database by biblioshiny," *Sci. Edit.*, vol. 8, no. 1, pp. 72–78, 2021.

[10] H. Ejaz, H. M. Zeeshan, F. Ahmad, S. N. A. Bukhari, N. Anwar, A. Alanazi, A. Sadiq, K. Junaid, M. Atif, K. O. A. Abosalif, et al., "Bibliometric analysis of publications on the omicron variant from 2020 to 2022 in the scopus database using r and vosviewer," *Int. J. Environ. Res. Public Health*, vol. 19, no. 19, p. 12407, 2022.

[11] J. Wang, X. Li, P. Wang, and Q. Liu, "Bibliometric analysis of digital twin literature: A review of influencing factors and conceptual structure," *Technol. Anal. Strateg. Manage.*, vol. 36, no. 1, pp. 166–180, 2024.

[12] Y. Yu, Y. Li, Z. Zhang, Z. Gu, H. Zhong, Q. Zha, L. Yang, C. Zhu, and E. Chen, "A bibliometric analysis using vosviewer of publications on covid-19," *Ann. Transl. Med.*, vol. 8, no. 13, 2020.

[13] S. Harnal, G. Sharma, S. Malik, G. Kaur, S. Khurana, P. Kaur, S. Simaiya, and D. Bagga, "Bibliometric mapping of trends, applications and challenges of artificial intelligence in smart cities," *EAI Endorsed Trans. Scalable Inf. Syst.*, vol. 9, no. 4, p. 8–8, 2022.

[14] S. R. Nallola and V. Ayyasamy, "Analyzing real-time surveillance video analytics: A comprehensive bibliometric study," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 3, pp. 345–356, 2023.

[15] W. Scheibel, D. Limberger, and J. Döllner, "Survey of treemap layout algorithms," in *Proc. 13th Int. Symp. Visual Inf. Commun. Interact.*, 2020, pp. 1–9.

[16] G. Kaur, "Bibliometric mapping of theme and trends of blockchain," in *2024 11th Int. Conf. Reliability, Infocom Technol. Optimization (Trends Future Directions) (ICRITO)*, 2024, pp. 1–6.

[17] O. Sharma, S. Mohapatra, J. Mohanty, P. Dhiman, and A. Bonkra, "Predicting agriculture leaf diseases (potato): An automated approach using hyper-parameter tuning and deep learning," in *2023 Third Int. Conf. Secure Cyber Comput. Commun. (ICSCCC)*, 2023, pp. 490–493.

[18] J. Perez-Martin, B. Bustos, S. J. F. Guimaraes, I. Sipiran, J. Pérez, and G. C. Said, "A comprehensive review of the video-to-text problem," *Artif. Intell. Rev.*, pp. 1–75, 2022.

[19] M. A. Rojas-Sánchez, P. R. Palos-Sánchez, and J. A. Folgado-Fernández, "Systematic literature review and bibliometric analysis on virtual reality and education," *Educ. Inf. Technol.*, vol. 28, no. 1, pp. 155–192, 2023.

[20] P. Kaur, A. Kaur, M. Khurana, and R. Damaševičius, "Multimodal hinglish tweet dataset for deep pragmatic analysis," *Data*, vol. 9, no. 2, p. 38, 2024.

[21] A. Kaur and M. Khurana, "Multimodal Sentiments: Unraveling Text and Emoji Dynamics Through Deep Learning," in *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, pp. 1-6, March 2024.