

Received on: 24 May 2026

Revised on: 30 June 2026

Accepted on: 15 July 2026

Generative Artificial Intelligence in Journalism: A Bibliometric Analysis of Research Trends (2022-2026)

Mohd Anas

Assistant Professor, Aligarh Muslim University

anas.mc@amu.ac.in

Abstract

This study maps the development of research on generative AI and journalism between 2022 and 2026 using bibliometric analysis on publication data retrieved from the Scopus database. The data was analysed using the Bibliometrix package through the Biblioshiny interface in RStudio. The interface is capable of producing multiple publication trends and indicators such as leading sources, geographical distribution, keyword relationships, and thematic development etc. Using appropriate keywords and filters on the Scopus website a corpus of 1,148 publications was produced and analysed. The findings indicate substantial growth in research output, with an annual growth rate of 42.02%, demonstrating the rapid expansion of scholarly attention towards AI-driven journalism. The literature is primarily published in specialist journalism and communication journals, while contributions from multidisciplinary and technology-oriented sources highlight the interdisciplinary nature of the field. The United States emerged as the leading contributor, although research activity was distributed across multiple regions. Keyword analysis revealed that artificial intelligence, journalism, generative AI, ChatGPT, automated journalism, ethics, and disinformation represent central research themes. Thematic evolution indicates a shift from broad discussions of AI applications towards more focused investigations of generative AI, large language models, transparency, credibility, and responsible AI practices. This study provides a structured overview of the rapidly expanding literature on generative AI and journalism and it offers useful insights for researchers, journalism educators, media organisations, and policymakers.

Keywords: Generative artificial intelligence; Artificial intelligence; Journalism; ChatGPT; Large language models; Automated journalism; Bibliometric analysis; Bibliometrix; Biblioshiny

Introduction

Artificial intelligence (AI) has become one of the most influential technological developments of recent times. The rise of applications like ChatGPT built around LLMs and its rapid adoption has resulted in multifarious discussions and debates in the media and the public sphere. It has fundamentally altered both industry practices and the trajectory of academic

research, resulting in an exponential rise in scholarly literature. While scholarly publications, specifically focussing on the science and engineering part have increased substantially since 2022 (Farhat et al., 2023), existing research is dispersed across computer science, information science, communication and related disciplines.

In Journalism, an increasing number of studies has explored various aspects of generative AI, but the current literature remains fragmented and there is limited understanding of the field's overall thematic structure and orientation. To address this gap, this study provides a bibliometric analysis of literature on artificial intelligence, generative AI and journalism published between 2022 and 2026 from the Scopus database. 2022 has been chosen as the starting year because it was in this year that ChatGPT was released.

Bibliometric analysis offers a systematic approach for examining the development of a research field by analysing publication patterns, citation relationships, collaboration networks, and thematic structure (Passas, 2024). The analysis is helpful in providing an overview of how knowledge in a discipline has evolved over time, identify the principal directions of scholarly activity enabling to "unpack the evolutionary nuances of a specific field, while shedding light on the emerging areas in that field" (Donthu et al., 2021, p. 285). In this study the bibliometric analysis was done through Biblioshiny which is the web-based graphical interface of the Bibliometrix software package developed by Aria and Cuccurullo (2017) and is considered an excellent tool for producing bibliometric indicators. While there are multiple softwares like VOSviewer, CiteSpace etc., Biblioshiny was selected because of its wide range of bibliometric indicators and visualization capabilities.

Methodology

The study employs a quantitative approach on publication data retrieved from the Scopus database. Scopus was chosen because it is the "world's largest abstract and citation database of scientific literature" (Schotten, 2017), moreover, it provides extensive coverage of communication, journalism, media studies, and computer science journals. The following search string was used:

TITLE-ABS-KEY (("artificial intelligence" OR "generative AI" OR ChatGPT OR GPT OR "large language model*" OR LLM* OR "foundation model*" OR "machine learning" OR "automated journalism" OR "algorithmic journalism") AND (journalism* OR newsroom* OR reporter* OR editor* OR "news production" OR "news writing" OR "news generation" OR "news media" OR "news organization*" OR "AI-assisted journalism" OR "computational journalism" OR "robot journalism")). The string was further refined through applying the following filters: Subject Areas (Computer Science, Arts and Humanities, Decision Sciences, Multidisciplinary, and Social Sciences), Document Type (Article), Language (English), and relevant author keywords associated with artificial intelligence and journalism.

The TITLE-ABS-KEY search field retrieves records containing the specified search terms in the article title, abstract, or keywords. The time period was defined from 2022-2026 and only published articles were part of the search. All subjects related to life sciences, engineering

and health fields were excluded. Social sciences, computer science, arts and humanities and multidisciplinary were part of the search string . The search produced 1,148, documents after applying the selected filters and formed the dataset used for the bibliometric analysis. The results were exported as CSV from the Scopus website and were imported into Biblioshiny. A verification was done to check the completeness of metadata then various descriptive bibliometric indicators and data was generated.

The following workflow was utilised i: 1) R and RStudio were installed. 2) The Bibliometrix package was installed in R. 3) Biblioshiny was launched through RStudio using the Bibliometrix package. 4) Scopus data in the form of CSV file were imported into Biblioshiny for bibliometric analysis. The software was then used to generate bibliometric indicators such as descriptive statistics, annual scientific production leading publication sources, geographic sources of the research, Keyword Co-occurrence Network and thematic maps.

Results

Descriptive Analysis

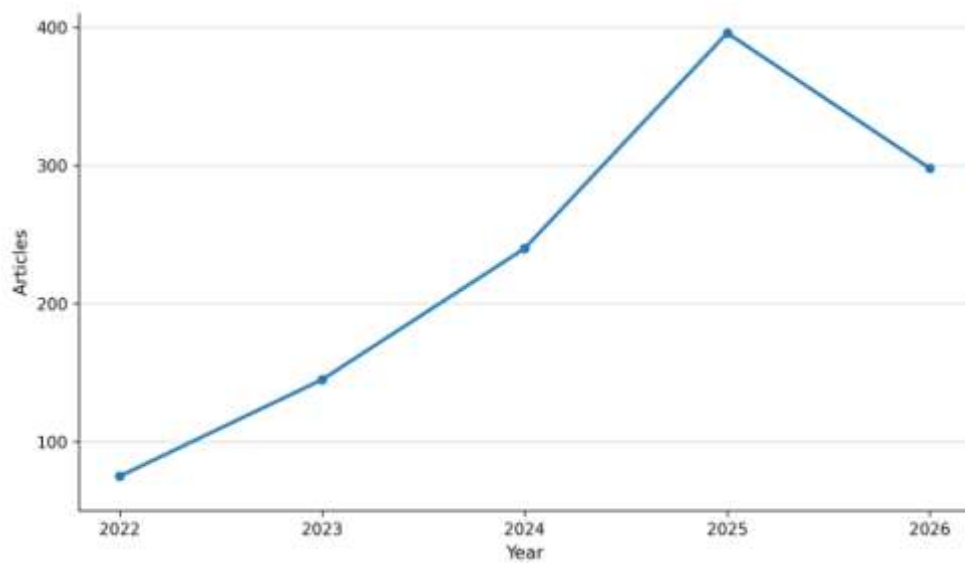
Table 1 Descriptive Analysis

Indicator	Value
Timespan	2022-2026
Documents	1,148
Sources	453
Authors	3,169
References	54,178
Average citations per document	11.40
Annual growth rate	42.02%
International co-authorship	28.22%

The dataset was made up of 1148 journal articles published between 2022 and 2026. These publications involved 453 sources, 3,169 authors 54,178 references. The annual growth rate of 42.02% indicates a rapidly expanding body of literature during the study period. The average number of citations per document was 11.40, while international co-authorship accounted for 28.22% of the publications, reflecting a moderate level of international research collaboration.

Annual Scientific Production

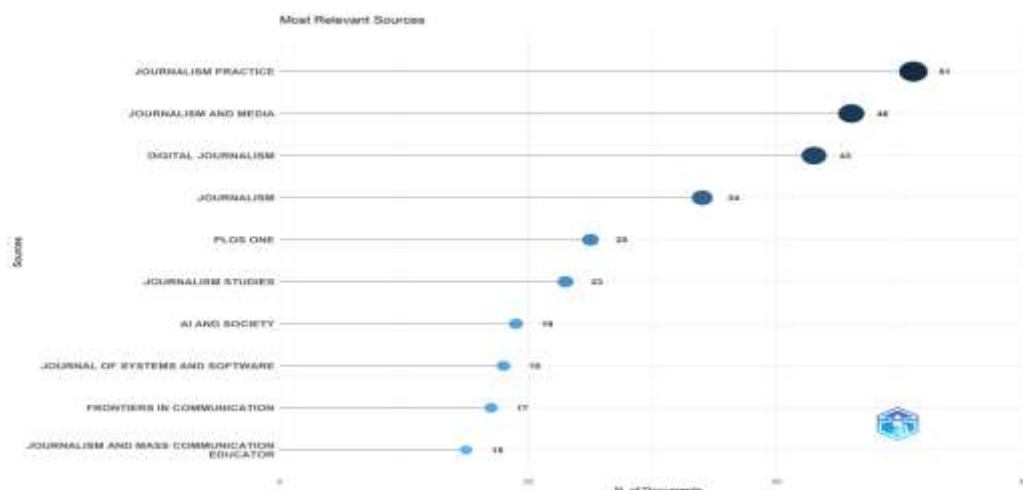
Figure 1: Annual scientific production on generative artificial intelligence and journalism published in Scopus between 2022 and 2026. The lower publication count for 2026 reflects an incomplete publication year.



In the time period of this study, the research output increased steadily and an accelerated growth can be observed after 2023, since ChatGPT was released in 2022. The largest number of articles was published in 2025, and the apparent decline in 2026 reflects an incomplete publication year.

Leading Publication Sources

Figure 2: Most relevant publication sources for research on generative artificial intelligence and journalism (2022-2026).



The leading publication sources data suggests that research is concentrated in a relatively small group of journals with specialist journalism and media journals like *Journalism Practice*, *Journalism and Media* and *Digital Journalism*. The presence of interdisciplinary journals suggests collaboration between journalism, communication, information science, and

computer science. The distribution also indicates that AI and journalism has become an established research topic within communication studies. Finally, the discussion of generative AI in the Scopus database has been led primarily by journalism scholars rather than by computer science researchers.

Geographical Distribution of Research

Table 2 Top 10 countries by scientific production in research on generative artificial intelligence in journalism (2022-2026).

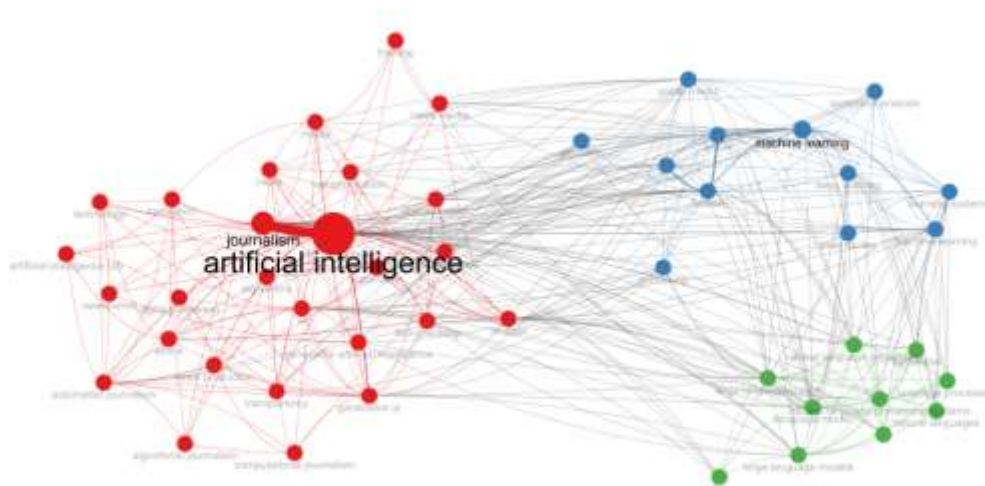
Rank	Country	Publications
1	USA	541
2	Spain	194
3	China	187
4	United Kingdom	160
5	Germany	149
6	Australia	101
7	India	99
8	Netherlands	77
9	Italy	51
10	Brazil	49

Research output is concentrated in a relatively small group of countries, with the USA being the leading contributor followed by Spain. English-speaking countries (USA, UK, Australia) and several European countries (Spain, Germany, Netherlands, Italy) are strongly represented. India appears in the top 10, indicating growing participation in this field.

Keyword Co-occurrence Network

Figure 3: Keyword co-occurrence network based on author keywords illustrating thematic relationships and research clusters in the literature on generative artificial intelligence in journalism (2022-2026).

The keyword co-occurrence network (KCN) is a graphical representation of the author keywords which appear together frequently across research papers. KCN provides insight into the intellectual structure of the field; Radhakrishnan et al., 2017 explain the knowledge mapping potential of KCN; here "each keyword is represented as a node and each co-occurrence of a pair of words is represented as a link. The number of times that a pair of words co-occurs in multiple articles constitutes the weight of the link connecting the pair" (para 1).

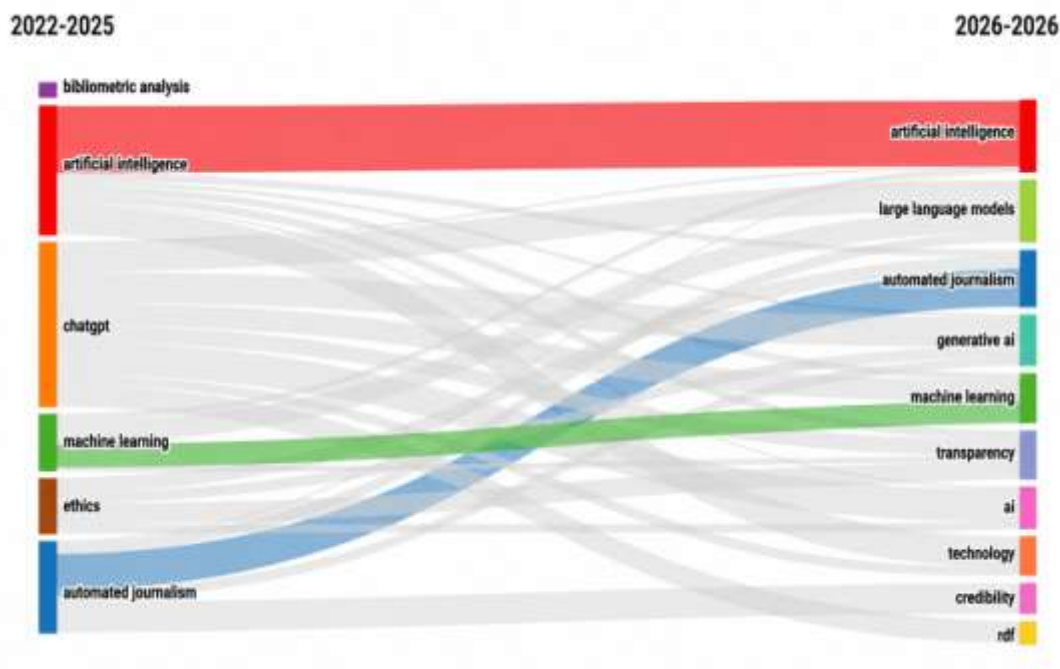


The KCN in this study identified three major thematic clusters. Artificial intelligence, the largest and the most dominant one is represented by red colour. It is closely linked to themes like generative AI, ChatGPT, journalism, misinformation, ethics, news media, social media, and automated journalism, suggesting that these topics form the core of current scholarly discussions. The green cluster represents publications focused on machine learning computational approaches and related themes. The high internal density of the network suggests these technologies are a major methodological dimension in journalism research. The blue cluster brings together keywords related to technology such as algorithm, machine learning and deep learning. It is distinct from the other clusters but still well connected, indicating the integration of technological advances within journalism studies. All three clusters are linked by numerous connections, signifying a high degree of interconnectedness among research topics, with artificial intelligence acting as the central bridge connecting diverse themes. The relative size of the node representing artificial intelligence indicates that it as the most frequent author keyword in the data set. Other large nodes are journalism, ChatGPT, generative AI, and machine learning, highlight their growing significance in recent publications.

Thematic Evolution

Figure 4: Thematic evolution of author keywords in artificial intelligence and journalism research (2022-2026).

Figure 4 illustrates the evolution of major research themes between the two time periods (2022-2025 and 2026). Artificial intelligence remained the dominant theme throughout the study period. Early interest in ChatGPT diversified into related themes such as large language models, generative AI, and broader AI applications, reflecting increasing specialization of research.

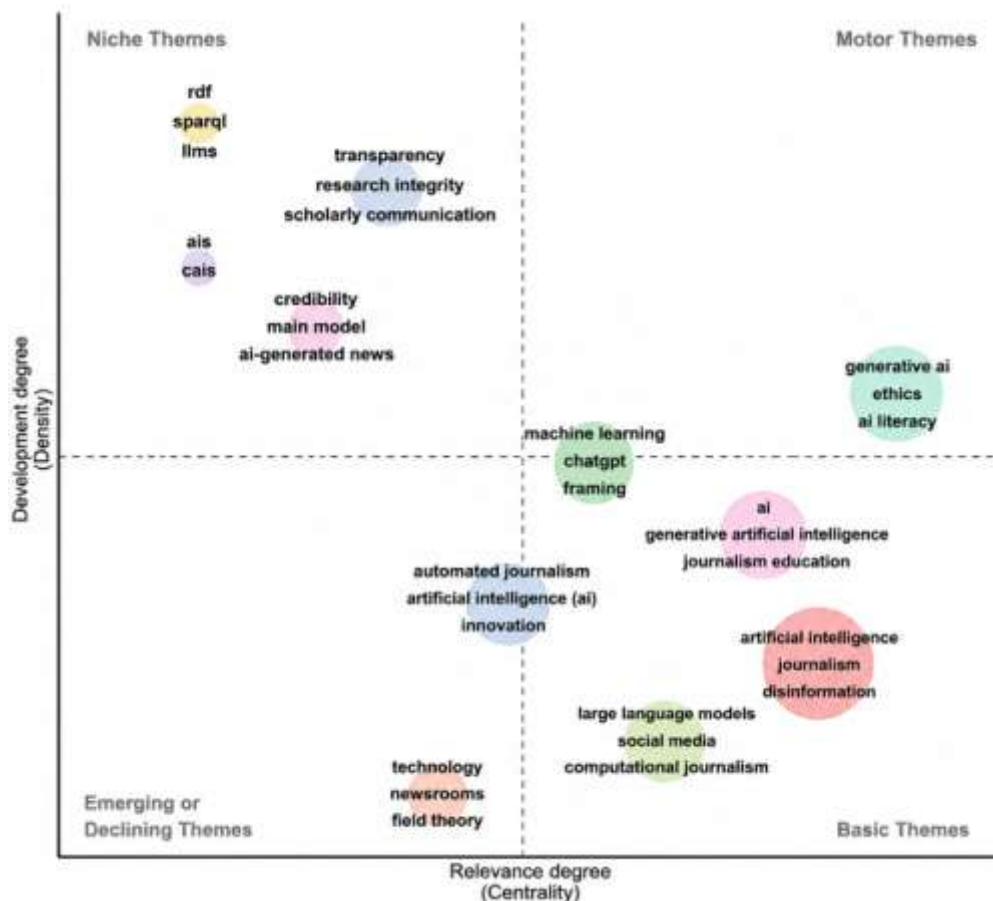


The persistence of automated journalism across both period is another major observation, indicating that it has become an established area of inquiry; this finding is also supported by industry practice where news organisation use algorithmic systems to generate data-driven stories, corporate financial reports and sports news summaries (Diakopoulos, 2019; Graefe, 2016). Machine learning continued as an important technological foundation, remaining connected with several emerging research directions. New themes including transparency, credibility, and technology became more visible in the later period, suggesting growing attention to the societal and ethical implications of AI in journalism.

Thematic Map (2026)

Figure 5: Thematic map based on author keywords illustrating the conceptual structure of research on generative artificial intelligence in journalism during 2026.

Biblioshiny produces a thematic map with four quadrants which can classify research themes according to centrality and density. The centrality depicts the important of themes within the research field; whereas density shows the level of internal development. In the basic themes quadrant, artificial intelligence, journalism, and disinformation are dominant themes indicating that these topics are central to the field provide the conceptual foundation for research. The motor themes quadrant includes machine learning, natural language processing, and social media, these themes have high centrality and density indicating that these are well-developed themes that drive current research. Automated journalism and computational journalism appear within the niche themes quadrant, indicating their relative specialisation within the broader literature.



The map shows that technological themes are increasingly linked with journalism research, demonstrating the integration of artificial intelligence methods with media practice, news production, and communication studies. The overall thematic structure indicates that research has progressed beyond general discussions of artificial intelligence toward more specialized investigations of generative AI, machine learning, natural language processing, and automated journalism.

Discussion

The findings show that research on artificial intelligence and its various aspects in journalism has expanded rapidly since 2022 and has become a major topic of scholarly interest. Across the 5-year timespan chosen in this study (2022-2026) the field demonstrated an exceptionally high growth rate of 42.02 % (Table 1); suggesting a hyper-rapid academic response to the AI boom. The presence of 54,178 references in the selected documents further suggest that the discipline is pulling from outside fields and is super- interdisciplinary in nature. While journalism-artificial intelligence interface was already explored through topics like automated news writing and data-driven reporting (Bartleman & Schapals, 2026); the sheer amount of media attention generated after the release of ChatGPT in 2022, and the immense possibilities of generative AI has attracted researchers from diverse disciplines such as journalism, communication, computer science, and related disciplines.

The identification of leading publication sources such as *Journalism Practice*, *Journalism and Media*, *Digital Journalism*, and *Journalism Studies* (Figure 2), suggests that AI has become firmly embedded within journalism scholarship. The presence of multidisciplinary journals such as *PLOS ONE* and *AI and Society* further reflect the interdisciplinary character of the field

The geographical distribution of publications is not surprising. Research activity is concentrated within a relatively small group of countries, already known for prodigious research output. The United States contributed the largest number of publications, followed by Spain, China, the United Kingdom, and Germany. However, the presence of India among the most productive countries suggests that research on artificial intelligence and journalism is expanding beyond its traditional centres in the western hemisphere.

In the KCN generated in this study (Figure 3), artificial intelligence occupies the central position within the network and is closely linked with journalism, generative AI, ChatGPT, ethics, misinformation, automated journalism, and social media. The dense relationships between artificial intelligence and its associates indicate that current research examines artificial intelligence alongside questions of journalistic practice, media systems, and public communication. The prominence of ethical issues and misinformation also suggests that researchers are concerned not only with technological capability but also with the consequences of AI adoption within news production.

A thematic map is also shown in Figure 5 to complement these results. It produces motor themes (top-right in Figure 5), basic themes (bottom-right in Figure 5), niche themes (top-left in Figure 5) and emerging or declining themes (bottom-left in Figure 5). Motor themes represent the topics central to a discipline, they drive the current research agenda, which in this case are generative AI, ethics and AI literacy, while machine learning and ChatGPT are close to the boundary between motor and basic themes. Artificial intelligence, journalism, and disinformation have high centrality but lower density, and they represent the basic themes. Specialised themes like transparency, research integrity, scholarly communication and ai-generated news, credibility etc. are present in the niche themes. In the emerging or declining themes quadrant, technology, newsrooms, and field theory are found. These are likely *emerging* conceptual frameworks that are just beginning to find their footing in the literature.

The thematic evolution analysis (Figure 4) indicates that the field witnessed rapid evolution within a short period of time. Early studies were primarily focused on artificial intelligence and its role in journalism. As the research grew, the focus shifted to more specific topics such as generative AI, large language models, transparency, credibility and responsible AI, suggesting that the literature is moving from broad exploratory discussions to more focused investigations of how generative AI influences journalistic work, newsroom practices and relationships between news organisations and audiences. Artificial intelligence and automated journalism are themes that are present in both time periods. Their continued

presence indicates that earlier research was the basis of current work. These results suggest that research on generative artificial intelligence in journalism is in a phase of rapid development characterized by growing thematic diversity and increased linkages between technological innovation and journalism studies. While generative AI has been the most dominant theme in recent years, the literature continues to wrestle with overarching themes such as ethics, misinformation, transparency, and newsroom practice. This balance suggests the field is moving beyond conversations about technological novelty and toward a more holistic exploration of how artificial intelligence is transforming journalism as a profession, a public institution, and a subject of academic inquiry.

Conclusion

This study mapped the development of research on generative artificial intelligence in journalism between 2022 and 2026 using bibliometric indicators produced through Biblioshiny software package. The findings show that the field has expanded rapidly within a relatively short period, with a 42.02 % annual rate observed in 1148 publications spread across 3,169 authors. The findings indicate that generative AI has become a major subject of academic inquiry in the corpus and is not viewed merely as an experimental disruption but as a permanent component of journalism and media production. The conceptual structure of the field has also evolved significantly. The field has matured from its early focus on ChatGPT toward more stable, "motor" themes, such as generative AI, ethics, and AI literacy. These topics now drive the research agenda, suggesting that scholars have moved beyond simply identifying AI tools and are now prioritizing frameworks for their responsible integration into newsrooms. While the United States remains the dominant hub for this research, the inclusion of countries like India and Brazil in the top 10 indicates that the scholarly investigation of AI in journalism is a global journalistic priority.

Limitations of the study and future avenues of research

The study has several limitations; first it is based only on the Scopus database and includes journal articles published between 2022 and 2026. Second, it is important to acknowledge that bibliometric analysis examines publication metadata and does not focus on detailed content or quality of individual studies (Donthu et al., 2021). The publication count for 2026 should also be interpreted with caution because it represents records retrieved up to the date of data collection. Future research may include more databases such as Web of Science, comparing publication patterns across databases, or it may combine bibliometric analysis with systematic or qualitative literature reviews.

References

1. Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959-975.
2. Bartleman, M., Schapals, A. K., & Dubois, E. (2026). Generative AI and the New Landscape of Automated Journalism: A Systematized Review of 185 Studies (2012–2024). *Journalism and Media*, 7(1), 39. <https://doi.org/10.3390/journalmedia7010039>
3. Diakopoulos, N. (2019). *Automating the News: How Algorithms Are Rewriting the Media*. Harvard University Press. <http://www.jstor.org/stable/j.ctv24w634d>
4. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285-296. <https://doi.org/https://doi.org/10.1016/j.jbusres.2021.04.070>
5. Farhat, F., Silva, E. S., Hassani, H., Madsen, D. Ø., Sohail, S. S., Himeur, Y., Alam, M. A., & Zafar, A. (2024). The scholarly footprint of ChatGPT: a bibliometric analysis of the early outbreak phase [Systematic Review]. *Frontiers in Artificial Intelligence, Volume 6 - 2023*. <https://doi.org/10.3389/frai.2023.1270749>
6. Graefe, A. (2016). *Guide to automated journalism*. Tow Center for Digital Journalism, Columbia Journalism School.
7. Passas, I. (2024). Bibliometric analysis: the main steps. *Encyclopedia*, 4(2).
8. Radhakrishnan, S., Erbis, S., Isaacs, J. A., & Kamarthi, S. (2017). Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. *PloS one*, 12(3), e0172778. <https://doi.org/10.1371/journal.pone.0185771>
9. Schotten, M., Meester, W. J., Steinginga, S., & Ross, C. A. (2017). *A brief history of Scopus: The world's largest abstract and citation database of scientific literature*. In Research analytics (pp. 31-58). Auerbach Publications.