A Scientometric Review of Automated Journalism: Analysis and Visualization

Zhichao Xu1,a, Xintong Lan1,b*

1Sun Yat-sen University Nanfang College School of Literatures and Communications Guangzhou, Guangdong
2zhichao.x@outlook.com
b* Corresponding author: lanxt@nfu.edu.cn

Abstract—In recent years, automated news has been more and more adopted by media organizations, which has aroused the attention and discussion in the industry and academia. However, few studies have attempted to map out the global research on automated journalism. To analyze the state of research evolution in automated journalism and its trends, this paper aims to analyze the year of publication, most participating countries, institutions, journals, and topics of the relevant papers collected from the Web of Science (WoS) database. The results show that there are many countries actively participating in automated-journalism-related research, and the USA is the most productive country followed by Germany and Spain. Most of the active journals are in the field of journalism and communication. However, the visualization map of author’s keywords implies the nature of interdisciplinarity of the research, covering the main disciplines of journalism, computer science and automation.

1. INTRODUCTION
Automated journalism, defined as the auto generation of journalistic stories with the use of software and algorithms, with no human intervention outside the initial programming [1]. The core technology of automated journalism is to sort, classify, correlate and filter massive data through algorithms, and adapt the collated data into the article template. With the development of data processing technology and natural language processing technology, algorithmic writing is now more and more widely used in the production of news content. In 2006 Thomson Financial, an American information provider, used a computer program to automatically write economic and financial news stories. In 2010, The Los Angeles Times took the lead in the adoption of computer program to automatically generate contributions in the Homicide Report. In 2014, the Associated Press began to use the WordSmith platform to automate its quarterly reporting, and it became the first world-class news outlet to use artificial intelligence writing. At present, many of the world’s leading news organizations, such as Forbes, The New York Times and Xinhua News Agency, are also trying to apply artificial intelligence technology to news production.

To enhance the adoption of automated journalism, various researches have been conducted in the last decade. Some researchers focus on the technical issues, while others deal with the non-technical issues such as the development, the limitations and the ethical issues of automated journalism. For example, zhang et al. investigated the possibility to automatically generate sports news from live text commentary scripts[2]. Graefe et al. conducted an online experiment to study people’s perception of automated computer-written news and they find that subjects rated articles declared as human written always more favorably and subjects rated computer-written articles as more credible and higher in journalistic expertise but less readable[3]. The Tow Center for Digital Journalism has launched a project called Covering COVID 19 with Automated Journalism, which study on how media organizations should leverage technological innovation in times of crisis[4].

Published under licence by IOP Publishing Ltd
Some literature review of automated journalism did exist[5-6], but limited efforts have been made to outline and visualize the research trends. This study attempts to conduct a scientometric review of the scientific literature relating to automated journalism and map out the annual trends, main countries, institutions, publications, and topics of the related research based on the WoS database. These maps will help researchers to better understand the current state of the automated journalism research in the world and identify the hot topics in the literature.

2. DATA AND METHODS
A WoS is one of the bibliometric database sources for searching the literature, so it was employed in this study[7]. The data were retrieved using “Automated Journalism”, “Robot journalism” and “Computer-generated News” as main keywords. The search strategy in retrieving publication during 1990–2020 time frame is as follows: TS = (“Automated Journalism” OR “Robot journalism” OR “Computer-generated News”) AND DOCUMENT TYPES: (All document types). Indexes: SCI-EXPANDED, SSCI, A&HCI, ESCI.

A total of 147 articles retrieved from the WoS database was analysed using the VOSviewer software. It can be helped in constructing and visualizing bibliometric networks, which include main countries, publications, citations, institutions, journals and authors keywords.

3. RESULTS FINDINGS
To explore the latest development of research on automated journalism and propose a reference for further study, the number of publications, main institutions, publications, disciplines, and keywords are presented as follows.

3.1 Annual Trends
Figure 2 shows the annual trends of publications during 2000-2020. It is clear that the number of publications form 2000 to 2012 are less than 5. However, from 2012, there was a remarkable increase in the total number of publications. Figure 3 shows that the increasing trend of the number of citations is consistent with the increasing trend of publications, thereby, reaffirms the increase awareness on automated journalism.
3.2 Main Countries and institutions
Table 1 shows the total number of publications (TP) from countries with a minimum of 5 publications on automated journalism. Based on the TP, the USA (TP, 37) tops the chart, followed by Germany (TP, 22) and Spain with TP of 17, and they have made up for about 50% of the 147 articles. The United States is recognized as the first country to begin the practice of automated news production. In March 2006, Thomson Corporation, an American information provider, used a computer program to realize the automatic generation of economic and financial news. Since then, automated journalism has been widely used in the production of news content in American news media. From Table 1, we can also see that most of the countries with high publications are from North America and Europe, which might be related to the development of natural language processing technology.

| Countries/Regions   | No. | % of 147 |
|---------------------|-----|---------|
| USA                 | 37  | 25.17%  |
| GERMANY             | 22  | 14.97%  |
| SPAIN               | 17  | 11.56%  |
| NETHERLANDS        | 10  | 6.80%   |
| ENGLAND             | 9   | 6.12%   |
| SWITZERLAND         | 7   | 4.76%   |
| BRAZIL              | 6   | 4.08%   |
| NORWAY              | 6   | 4.08%   |
| RUSSIA              | 6   | 4.08%   |

![Figure 2. Trends in publications from 2000 to 2020](image)

![Figure 3. Variation trend of citations from 2012 to 2020](image)
To further illustrate how those countries in Table 1 constitute research fronts, Figure 4 visualizes the bibliographic coupling networks of the institutions. It shows clearly that four major organizational clusters, namely the University of Amsterdam in Netherlands, University of Munster in German, University of Zurich in Switzerland, University of Oregon in the United States. The institutions with the same color (within a cluster) are more connected in terms of the research field and collaboration than those in another cluster.

![Bibliographic coupling of institutions](image)

Figure 4. Bibliographic coupling of institutions

3.3 Journals bibliometric participation

The journal that published articles on automated journalism were conducted in order to determine the most active journal in terms of a number of publications and link with other journals. Figure 5 illustrates that the main central cluster include journals such as Digital Journalism, Media Communication, Journalism Studies, Journalism practice, Professional De La Informacion, and Journalism and Mass Communication. Most of the journals are in the field of journalism and communication, and only one, the Computers in Human Behavior, is in the field of computer science. This implies that journals in the field of journalism and communication pay more attention to the research of automated news, for the development of robot writing technology do bringing great influence to the production of news content.
3.4 Co-occurrence of author’s keywords

The analysis of keywords in scientific publications is of great importance in knowing the research trends and their follow-up. In Figure 6, these keywords were grouped into 5 colored clusters based on similarities in the area of research. The first cluster (in blue) of automated journalism contains some technical topics such as artificial intelligence, chatbots and human-machine communication, with discussion on journalism studies and public service media. The second cluster, “journalism” in red, appears to be related to big data, data journalism and twitter, and it also related to some technical topics such as technology and natural language generation. The purple “robot journalism” cluster, which has a strong links with the “automated journalism” cluster, consists of topics such as computational journalism, algorithmic journalism and automated news. The green “automation” cluster contains the topic of bots, news production and online journalism. The fifth cluster of algorithms (in yellow), which relates to topics such as news, social media and credibility, has a strong link with other clusters. The map of author keywords indicates the nature of interdisciplinarity of the research, covering the related main disciplines of journalism, computer science and automation.

Figure 5. Bibliographic coupling of journals

Figure 6. A map based on author keyword co-occurrence network: clustered outcomes
4. CONCLUSION
The development of automated journalism, though still in its early stages, is of great importance for the future of news production. The essence of this study was to show the status and trends of research in automated journalism and propose a reference for further study. The findings have shown the following development on the research: (1) there has been a significant increase in publications and citations since 2012, and the USA is the most productive country followed by Germany and then Spain based on criteria for the number of publications; (2) main institutions include the University of Amsterdam, the University of Munster, the University of Zurich, the University of Oregon and the University of Vienna, which also reveals that Europe and the USA is leading in the research field; (3) the most active journals include Digital Journalism, Media Communication, Journalism Studies and so on, and most of them are in the field of journalism and communication; (4) the network of author’s keywords implies the nature of interdisciplinarity of the research on automated journalism, which focus on (a) journalism and technology, (b) artificial intelligence, (c) automation, and (d) algorithms. Automated journalism is still in its early stages and there are many technical issues need to be addressed. Future studies on automated journalism, on one hand, should keep focusing on its influence on the practice and concept of news, on the other hand, should pay more attention to its key technical problems. Only based on the strengthening of interdisciplinary exchanges can the research of automated journalism be continuously developed and promoted.

REFERENCES
[1] Andreas Graefe. Guide to Automated Journalism.Tow Center for Digital Journalism [EB/OL]. http://towcenter.org/research/guide-to-automated-journalism/,2016.
[2] Zhang J , Yao J G , Wan X . Towards Constructing Sports News from Live Text Commentary[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2016.
[3] Graefe A , Haim M , Haarmann B , et al. Readers’ perception of computer-generated news: Credibility, expertise, and readability[J]. Journalism, 2016:1464884916641269.
[4] Samuel Danzon-Chambaud, Covering COVID-19 with automated journalism: leveraging technological innovation in times of crisis. https://towcenter.columbia.edu/content/research-projects-0
[5] Jiang Zhongbo, Shi Xuemei. The Progress and Thinking of Algorithmic-Journalism-Related Research Abroad. Journalism And Mass Communication Monthly .06(2019):87-100. doi:10.15897/j.cnki.cn51-1046/g2.2019.06.010.
[6] WANG Yong, WANG Lei. Advantages, Disadvantages and Impacts: A Review on Robot Journalism Research. Journal Of Kunming University Of Science And Technology(Social Science) 17.06(2017):92-100. doi:10.16112/j.cnki.53-1160/c.2017.06.013.
[7] Abrizah A, Zainab AN, Kiran K, Raj RG (2013) LIS journal scientific impact and subject categorization: a comparison between Web of Science and Scopus. Scientometrics (94):721–740.