A Bibliometric Analysis of Communication Research on Artificial Intelligence and Big Data

Yujin Zhou[2] Han-Teng Liao[1,2,*]

[1]*New Media Research Centre, Sun Yat-sen University Nanfang College, Guangzhou, Guangdong, 510970, China*
[2]*School of Literatures and Communications, Sun Yat-sen University Nanfang College, Guangzhou, Guangdong, 510970, China*
[*]*Corresponding author. Email: h.liao@oxon.org*

ABSTRACT

The development of artificial intelligence (AI), big data and other communication technologies have multiple implications for the consumption and production of media content. Communication researchers have recently discussed several methodological and thematic challenges, ranging from the ways in which they disrupt the news production and consumption, to the digital transformation of the media and content ecosystems. With the aim to provide a systematic overview of the latest development of communication research in relation to AI and Big Data, this paper has mapped out the main sources, disciplines, and keywords based on a bibliometric analysis of 685 articles collected from Web of Science database. The findings show that the main clusters are "Communication and Sociology", "Journalism", and "Information and Telecommunications", with the cluster of "Communication and Political Science" being in-between them. Also, the visualization of the author keywords indicates the main "Communication and Sociology" cluster includes the main research topics such as surveillance, algorithms, datafication, privacy, ethics, etc., and the "Journalism" cluster includes computational journalism, data journalism, automated journalism. Although future work is needed to provide detailed synthesis, the paper nonetheless provides a succinct summary for communication researchers and professionals to have an overview of the latest discussions.
Keywords: big data, artificial intelligence, social media, machine learning, privacy, politics

1. INTRODUCTION

Given the advancement of artificial intelligence (AI) and Big Data, the communication sector faces several challenges regarding the ecosystems of news and information. Methodological and thematic challenges have been discussed, for example, on data journalism[1], automation of news[2–4], content personalization systems[5], public opinion[6], research methods[7] and ethics[8], etc. Given the fact that AI and Big Data technologies can enhance human and social cognition, they can also misrepresent, distort and misinform. Indeed, AI-powered misinformation campaigns can provide new ways to conduct cyber attacks [9]. As more and more people get their information and news from online platforms and mobile apps, researchers and policymakers have become increasingly concerned about the implications of the digital transformation of news and communication[10], including the communication management and strategy to respond to a pandemic such as COVID-19.

It is essential then to have an overview of the communication research on artificial intelligence and big data. However, there is little systematic bibliometric research work on the topic. This paper aims to fill the gap by providing an up-to-date bibliometric analysis, summarizing the latest research fronts.

2. DATA AND METHODS

With the aim to capture the relevant literature from the field of communication, this article used the following advanced search query for the Web of Science (WoS) database:

- TS = ("artificial intelligence" OR "machine learning" OR "big data") AND (SU=Communication OR WC=Communication)

Our query included the topics (TS) such as artificial intelligence, machine learning and big data, and any communication research based on the research area (SU) and Web of Science Category (WC).

We collected 685 articles on February 5, 2020 (including SCI-EXPANDED, SSCI, A&HCI, ESCI.).

We used the VOSviewer for visualization and identified research fronts by using co-citation, bibliographic coupling, and keyword co-occurrence analysis[11,12]. We also developed Python scripts to explore the relationship between publication source clusters and author keywords.

The year of 2013 is found to be a turning point as evidenced by the publication trendline shown in Figure 1. Before the year of 2013, at most 7 articles are published per year. Since 2014, the number of publications increases sharply. For the year of 2019, 164 articles are published.
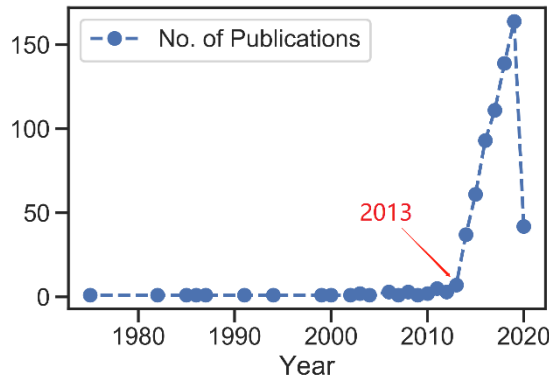
Figure 1 The number of publications

3. RESEARCH FINDINGS

The following sections summarize our findings on the significant research fronts, showing how main sources (and disciplines) are clustered first, and then how topics (i.e. keywords) are clustered.

3.1. Clusters of main sources

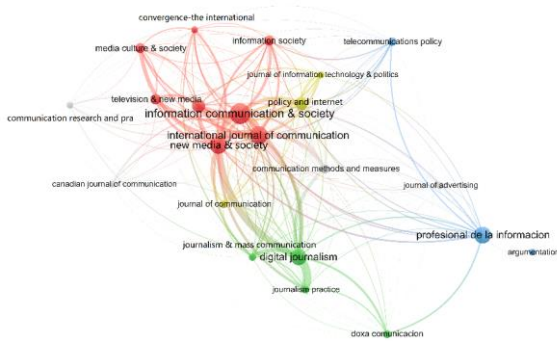Figure 2 and Figure 3 show how sources are clustered based on top publications and top-cited sources, respectively.

Figure 2 Top publication sources: a bibliographic coupling relationship network visualization
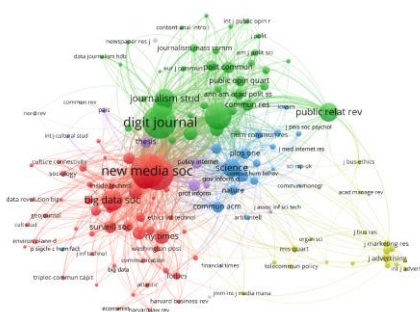

Figure 3 Top cited sources: a co-citation relationship network visualization

3.1.1. Clusters of publication sources

Figure 2 indicates where such work has been *published* and similar based on their bibliography (i.e. bibliographic coupling). The red "Communication and Sociology" cluster shows the top publication sources such as *Information Communication & Society*, *New Media & Society*, *Social Media + Society, Television & New Media*, *Media Culture & Society*, *Information Society*, etc. The cluster includes disciplines and research areas of sociology and communication. The green "Journalism" cluster contains mainly communication sources focusing on journalism such as *Digital Journalism, Journalism Practice, Journalism and Mass communication,* etc. The blue "Information and Telecommunications" cluster at the right includes sources such as *Profesional de la Informacion*, and *Telecommunications Policy*. Altogether, the three main clusters delineate the overall compositions of the publication landscape.

In-between these three main clusters, the yellow "Communication and Political Science" cluster is composed of *Policy and Internet, Journal of Information Technology & Polices,* and *Journal of Communication*. Though relatively smaller than the main three clusters, the "Communication and Political Science" cluster appears to play a bridging role among the three main clusters. Other journals such as *Communication Methods and Measures*, and *Journal of Advertising* also play similar bridging role.

3.1.2. Clusters of citied sources

Figure 3 shows where such work has been *citing* and similar based on their cited sources (i.e. co-citation).

The red "Communication and Society" cluster shows the mostly top publication sources such as *New Media Society*, *Media Culture & Society*, *Sociology*, etc., indicating the significant concerns of social implications.

The green "Communication and Journalism" cluster includes journals such as *Digit Journal, Communication Research* and *Public Relations Review*, focusing on behavioral science, political science, public relations, etc. It also contains diverse knowledge sources from political science, psychology, sociology and anthropology [13].

Also at the center of the cited work is the blue "Science and Computer Science". Although not as central, the yellow "Management, Business, Advertising and Marketing" cluster also connects the two main red and green clusters.

3.2. Main cluster of author keywords

Based on the author keywords, we identify 5 clusters as shown in Figure 4. At the centre, the green "big data" cluster includes keywords such as machine learning, privacy, sentiment, data mining, etc. Also, at the upper left corner, the red "AI" cluster features the central nodes of artificial intelligence and algorithms, along with terms such as communication, automation, computer, bots, agenda setting, etc.

Also, at the centre a bit to the bottom, the "media" cluster is shown to be related to power, surveillance, politics, social and critical. Different types of media include "new media", "mass media", "smart media", and "social media".

Figure 4 A map based on bibliographic keyword co-occurrence (author keywords)

In addition, at the upper centre, there are two clusters, the yellow "Internet and Journalism" cluster is situated at the upper centre, close to keywords such as transparency, content, knowledge, public, ethics. The blue-purple "digital advertising" cluster includes keywords such as advertising, personal, innovation, etc.

3.3. Relationship between sources and topics

The main "Communication and Sociology" cluster focuses more on topics such as privacy, ethics surveillance, algorithms, datafication, etc. For example, critical questions have been raised on big data[14]. The "Journalism" cluster focuses more on topics such as computational journalism, data journalism, automated journalism. For example, discussions have been summarized on the characteristics of agenda-setting on social media[15].

Altogether, the findings above can be summarized as follows: Most works are published in the three main clusters of "Communication and Sociology", "Journalism", and "Information and Telecommunications", with the cluster of "Communication and Political Science" being in-between the main three. Such work cites work beyond communication research, including sources such as *Science* and *ACM Communications*. While both AI and Big Data have become important topics, the main "Communication and Sociology" cluster has paid more attention to the social and ethical implications whereas the "Journalism" cluster has discussed more on computational journalism, data journalism, automated journalism, etc.

4. CONCLUSION

The above bibliometric analysis of 685 articles has shown how the current communications research is published, in relation to the main publication outlets, topics and disciplines. We have not only witnessed the rapid growth in publications since 2013, but also the two main clusters of research interests on (a) communication and society, and (b) journalism. The first cluster's focus on topics such as privacy, ethics surveillance, algorithms, datafication, etc., reveal the social implications of AI-powered news and communications systems, whereas the second main cluster's focus on topics such as computational journalism, data journalism, automated journalism points to the detailed discussions on the digital transformation of news production and consumption.

Our results indicate that communications research in relation to AI and big data can be summarized as the overall question of the digital transformation of societies dependent on communication and information systems that are increasingly automatized, personalized, and customized. Also, the findings also suggest that the discussions on big data have been more central when it comes to media, internet, news and politics. In contrast, the research on AI is in its early-development period.

Future work can potentially include in-depth analysis or expanded data collection of neighboring disciplines, such as information science, computer science and psychology. Also, the topics can be extended from misinformation to misbehavior online, especially such misbehavior is related to the use and misuse of information and content. Though our findings are preliminary, the systematic approach can be followed and improved by any future work on the topic.

ACKNOWLEDGMENT

REFERENCES

[1] R. Salaverría, Digital journalism: 25 years of research. Review article, El Profesional de La Información. 28 (2019). https://doi.org/10.3145/epi.2019.ene.01.

[2] S.C. Lewis, A.L. Guzman, T.R. Schmidt, Automation, Journalism, and Human–Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News, Digital Journalism. 7 (2019) 409–427. https://doi.org/10.1080/21670811.2019.1577147.

[3] M. Carlson, The Robotic Reporter: Automated journalism and the redefinition of labor, compositional

forms, and journalistic authority, Digital Journalism. 3 (2015) 416–431. https://doi.org/10.1080/21670811.2014.976412.

[4] M. Carlson, Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism, New Media Soc. 20.0 (2018) 1755. https://doi.org/10.1177/1461444817706684.

[5] B. Mittelstadt, Auditing for Transparency in Content Personalization Systems, International Journal of Communication. 10 (2016) 12.

[6] T. Gitlin, Media sociology, Theory and Society. 6 (1978) 205–253. https://doi.org/10.1007/BF01681751.

[7] J.W. Boumans, D. Trilling, Taking Stock of the Toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars, Digital Journalism. 4 (2016) 8–23. https://doi.org/10.1080/21670811.2015.1096598.

[8] M. Elish, danah boyd, Situating methods in the magic of Big Data and AI, Communication Monographs. (2017) 1–24. https://doi.org/10.1080/03637751.2017.1375130.

[9] A.S. Wilner, Cybersecurity and its discontents: Artificial intelligence, the Internet of Things, and digital misinformation, Int. J. 73.0 (2018) 308. https://doi.org/10.1177/0020702018782496.

[10]UN SECRETARY-GENERAL, Rise in Attacks on Media Must Not Become New Normal, Secretary-General Tells Geneva Correspondents' Association, Urging Bold Action to Defend Free Press, United Nations, 2019.

[11]E. Garfield, Research fronts, Current Comments. (1994).

[12]O. Persson, The intellectual base and research fronts of JASIS 1986–1990, Journal of the American Society for Information Science. 45 (1994) 31–38. https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<31::AID-ASI4>3.0.CO;2-G.

[13]J.N. Cappella, Vectors into the Future of Mass and Interpersonal Communication Research: Big Data, Social Media, and Computational Social Science, Human Communication Research. 43 (2017) 545–558. https://doi.org/10.1111/hcre.12114.

[14]D. Boyd, K. Crawford, Critical questions for big data: Provocations for a cultural, technological, and

scholarly phenomenon, Informacios Tarsadalom. 15 (2012) 662–679.

[15]H. Fu, X. Ren, H. Li, Z. Du, The roles and characteristics of social media for agenda-setting, in: 2014: pp. 1971–1976. https://doi.org/10.2495/MIIT132482.