Mapping Big Data and Artificial Intelligence in arts and humanities across time: an exploratory scientometric analysis based on the WoS database

Zhichao Xu1, Yuan Feng2, Han-Teng Liao1* and Xiaolin Jiang1

1 New Media Research Center, Sun Yat-sen University Nanfang College, Guangzhou, Guangdong, 510970, China
2 School of Communication and Design, Sun Yat-sen University, Guangzhou, Guangdong, 510275, China
*Corresponding author’s e-mail: h.liao@oxon.org

Abstract. The ways in which artificial intelligence (AI) and Big Data shape the work and the role of artists and humanities scholars have been at the forefront discussions among advanced universities and institutions, including examples such as computer-generated artwork, ethics, and philosophy, etc. With the aim to systematically review the related literature in arts and humanities, it is important to examine the discipline dynamics across time, thereby exploring both historical and current status. Based on the bibliographic data of 2,707 articles collected from Web of Science from 1956 to 2020, this paper has identified main disciplines and publications across three time periods, using various clustering and visualization methods. Based on the analytical and visualization findings, this article provides both a roadmap and data processes for artists, researchers, and policy-makers to understand how AI and Big Data have been discussed in arts and humanities disciplines.

1. Introduction
In recent years, the impact of Big Data and artificial intelligence (AI) applications on humanities, art, and philosophy have been discussed, especially about the promises and potential dangers of such applications. For instance, to ensure the development of responsible artificial intelligence, how to develop AI responsibly for human beings is a pressing issue for researchers and policy makers[1,2]. As another example, as forms of human expressions and societies are fundamental to the practices and theories of arts and humanities scholars, Yang Qian, the President of the 2019 International Joint Conference on Artificial Intelligence (IJCAI 2019), stated that "the second phase of AI will truly reshape human society, giving it its future form"[3]. Thus, the ways in which AI, along with the Big Data of human expressions and activities, reshape human and social forms of expressions become a central question facing arts and humanities scholars and practitioners. Advanced universities such as Stanford[4] and Oxford [5] have begun to develop humanities-related institutions surrounding AI, echoing the need to address various ethical implications and exploring the ways in which AI technology benefit humanity more broadly [1,3,6].

The discussions and applications of Big Data and AI in arts and humanities appears to be growing. For instance, philosophy of information has become fertile grounds for research and policies[7]. Digital humanities also play an important role[8]. For another, data visualization can provide new knowledge for understanding human experience, social relationships and network relevance[9]. AI also has begun
to reshape music expressions and compositions[10,11]. Nonetheless, few literature review papers have been published regarding the overall picture of the status of Big Data and AI in arts and humanities. With the purpose to fill such a gap, this paper aims to provide a systematic account of current status.

2. Data and methods
Following the conventions of scientometric analysis, the section describes the query design and methods.

2.1. Data sources
Often used for bibliometric analysis[12], the ISI Web of Science (WoS) is used to gather related literatures. In order to cover comprehensively arts and humanities research in relation to the topic of AI and Big Data (coded by the WoS field as TS), we have executed the Advanced Search query as below, covering 15 WoS research areas (coded by the WoS field tag as SU):

(TS = ( AI OR "Artificial Intelligence" OR "machine learning" OR "Big Data")) AND (SU=(“Arts & Humanities” OR ”Architecture” OR ”Art” OR ”Arts & Humanities Other Topics” OR ”Asian Studies” OR ”Classics” OR ”Dance” OR ”Film, Radio & Television” OR ”History” OR ”History & Philosophy of Science” OR ”Literature” OR ”Music” OR ”Philosophy” OR ”Religion” OR ”Theater”))

The bibliometric data was collected on 2019.11.04 with 2,707 articles.

2.2. Methods and Tools
To assist our analysis of the data, we first used VOSviewer to explore and analyse the network maps using journal publication sources as the main unit of analysis.

3. Research mapping results
This section presents the findings using a historical perspective.

3.1. Annual trends
Specific time periods have been identified based on the linear regression modelling of the number of articles (in log values) across time. As shown in Figure 1(a)-(d), the number of publications has continued to grow exponentially, except for the time period between 1986 and 2003.

Figure 1. Annual Publication Trend from 1956 to 2020.
Figure 1(e)-(h) shows the corresponding networks of publication outlets, highlighting the significant changes across time. Figure 1(e) shows that Voprosy Istorii as the top Russian humanities publication.
Figure 1(f) shows Zygon, a.k.a. the Journal of Religion and Science as the top. Figure 1(g)-(h) shows the Journal of New Music Research, which is a peer-reviewed academic journal covering research on musicology (including music theory), philosophy, psychology, acoustics, computer science, etc.

3.2. Top publication outlets
Table 1 lists the top publication outlets across three time periods, indicating significant shifts in main publication outlets and the dynamics of interdisciplinary collaboration.

| 1956-1985       | 1986-2003       | 2004-2020       |
|-----------------|-----------------|-----------------|
| Voprosyi Istori | Voprosyi Istori | Journal of New |
| American Historical | Arbor-Ciencia | Music Research |
| Review          | Pensamiento Y   |                 |
| Voprosy Filosofi| Daedalus        | Ethics and      |
| Istoriya Ssr    | Isis            | Information     |
| Deutsche Zeitschrift Fur Philosophie | American Historical Review | Technology |
|                 |                 | Leonardo        |
|                 |                 | Synthese        |
|                 |                 | Journal of      |
|                 |                 | Consciousness   |
|                 |                 | Studies         |

The publication outlets during the time period of 2004-2020 clearly differ from the other two. For instance, dedicating to the advancement between moral philosophy and information and communication technology (ICT), the journal Ethics and Information Technology has become the one of the main publication outlets after 2003. Also, covering the application of contemporary science and technology to the arts and music, Journal of New Music Research, and Leonardo have become important publication outlets after 2004, indicating strong research interest in the integration of science, technology, art, and humanities. Note that during the first time periods before 2003, there were several European continental publication outlets.

3.3. Publication clusters
Figure 2 shows the clustering outcomes based on Figure 1(g), revealing important outlets for each cluster.

![Figure 2. Clustering outcomes of publication outlets from 2004-2020.](image)

Four clusters have been identified and visualized: the ethics/philosophy cluster (in blue color, with Synthese at its core position), the arts/consciousness cluster (in orange), the music/technology cluster (in purple color, with Journal of New Music Research), and the ethics/technology cluster (in green color, with Ethics and Information Technology at its core position).
4. Conclusion
This literature review contributes to a better understanding of how humanities literature related to AI and Big Data change across time. By presenting findings across three time periods, the paper shows the mainstreaming of Big Data and AI applications, especially expressions of music. We also have found that, in the humanities field, cross-disciplinary collaboration is evident especially in research that intersects with science and technologies.

Bibliometric methods and regression modelling have helped to identify major time periods and publication outlets. Focusing on the most recent period of outlets from 2004-2020, based on the network analysis of the publication outlets, four clusters have been identified: the ethics/philosophy cluster, the arts/consciousness cluster, the music/technology cluster, and the ethics/technology cluster. Although we may miss the related research published in technology journals such as computer-generated graphics, the paper nonetheless provides important findings and methods for researchers, funding agencies, policymakers, and industry professionals to understand the dynamics in this area. Future research can build upon our research to advance such understanding.

Acknowledgments
The research is funded by the Application-oriented Curriculum Development Projects of “API, Machine Learning and Artificial Intelligence” (NFU 02-40248) and “Big Data” (NFU 02-40249), under the Guangdong Province Department of Education 2018 "Innovation-Strengthening Higher Education Program" Grants, and partly by a project of HCF for Education Development (2018WQNCX283), under the 2018 Major Projects of Guangdong Province Department of Education “Young Innovative Talents” Grants. Zhichao Xu, Yuan Feng and Han-Teng Liao have made equal contributions to the paper, including the data collection and analysis work and have approved the current version for publication.

References
[1] UNESCO (2018) Towards a global code of ethics for artificial intelligence research. UNESCO 2018: 3.
[2] Dignum, V. (2019) Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. (Springer International Publishing)
[3] Yang, Q., Wang, C. (2018) The second phase of AI will truly reshape human society, giving it its future form. UNESCO
[4] Stanford University (2019) Stewarding AI in the Age of Big Tech. Stanf. Inst. Hum.-Centered Artif. Intell.
[5] Choudhury, N. (2019) £150 million funding for AI research at University of Oxford. Open Access Gov.
[6] Boddington, P. (2017) Towards a Code of Ethics for Artificial Intelligence. (Springer International Publishing)
[7] Floridi, L. (2002) What is the Philosophy of Information? Metaphilosophy 33: 123–145.
[8] Sculley, D., Pasanek, B.M. (2008) Meaning and mining: the impact of implicit assumptions in data mining for the humanities. Lit Linguist Comput 23.0: 409.
[9] McCosker, A., Wilken, R. (2014) Rethinking ‘big data’as visual knowledge: the sublime and the diagrammatic in data visualisation. Vis. Stud. 29: 155–164.
[10] Widmer, G. (2002) Machine discoveries: A few simple, robust local expression principles. J. New Music Res. 31: 37–50.
[11] Moroni, A., Manzolli, J., Zuben, F.V., Gudwin, R. (2000) Vox populi: An interactive evolutionary system for algorithmic music composition. Leonardo Music J. 49–54.
[12] Cui, T., Zhang, J. (2018) Bibliometric and review of the research on circular economy through the evolution of Chinese public policy. Scientometrics. 1013–37.