Artificial Intelligence and Social Computing, Vol. 122, 2024, 315–324
https://doi.org/10.54941/ahfe1004668

A Systematic Review of Changing Conceptual to Practice AI Curation in Museums: Text Mining and Bibliometric Analysis

Shengzhao Yu[1], Jinghong Lin[2], Jung Huang[2], and Yuqi Zhan[2]

[1]School of Visual Art Design, Guangzhou Academy of Fine Arts, Guangzhou, China
[2]Department of Art and Sciences, Guangzhou Academy of Fine Arts, Guangzhou, China

ABSTRACT

The rapid development of artificial intelligence (AI) algorithms has accelerated the global digitization of museums. This study was conducted to clarify conceptual change to practice by applying a systematic literature review to a combination text mining and bibliometric analysis technique to visualization network. Based on the study selection articles from Web of Science (WOS). Our research questions focused on revealing the interconnected network of digital museum collections, expert knowledge and algorithms, and recommendation systems. The findings showed that 288 articles were finally selected to be analyzed.Conceptualizing AI curation in museums is currently underwayincombining AI with museum curatorial knowledge and innovate the practice mode of public participation in museum AI curation With emphasis on the the exchange of domain knowledge process. Moreover, three dimensions to consider including (1) design dimension focus on Methods and approaches for curating museum artificial intelligence exhibitions, (2) learning dimension focus on iterative development of new algorithm models guides the practice of intelligent curation, and (3) standard dimension focus on assessment and evaluation inpublic participation in curating museum cultural heritage exhibitions. In addition, the museum and AI community will mutually benefit. In particular, the convergence of new technologies and the exchange of domain knowledge would result in fairer and safer applications in the future as a result of learning from one another's flaws.

Keywords: Artificial intelligence, Algorithm, Cultural heritage, Museum collection, Curation

INTRODUCTION

Museums are non-profit public institutions that study, collect, preserve, interpret, and exhibit tangible and intangible cultural artifacts. To provide multiple educational, appreciative, reflective, and knowledge-sharing opportunities. Museum curation refers to the practice of selecting, organizing, and displaying objects and ideas in a museum setting, as well as interpreting cultural heritage materials, including the exploration of the stories and meanings behind them (Campbel, 2020). For the concept of contemporary curation, 1) curation is about organizing, managing, and combining information, data, concepts, images, artworks, and other objects; 2) curation is a form of knowledge production and power distribution; 3) curation occurs inside and outside of museums and galleries, and increasingly involves content and data management activities; 4) curation is consistently performed by using available technical tools and cross-social technical networks and is supported by philanthropic organizations; 5) An event open to a wider audience than just arts professionals (Tyzlik-Carver, 2016). Museums are entities and collecting institutions governed by curators. When seeking cultural democracy and challenging the traditional authority of museums in the digital context, museums are considering a variety of ways to become a contact zone where visitors and collections can continuously connect and form relationships. Curations can take various forms (Axelsson et al., 2022). Similar "curation" processes are carried out by a variety of actors, including traditional creators, public users, and computer algorithms (Thorson et al., 2016). Through forums, search engines, ML models, and ecosystems, museums, curators, and audiences have influenced the current flow of museum knowledge. Consequently, as museum collection data flows through the global computing space, the ability to interpret and construct historical knowledge in numerous ways has emerged in terms of curation. The production of immaterial information and knowledge has been altered by computers and intelligence knowledge (Diakopoulos, 2019). In the context of massive new social data and production processes driven by the emergence of new technologies, an increasing number of researchers have theorized the future of digital data, cultural heritage, curation, and human concepts (Cameron et al., 2021).

AI algorithms can support systematic and structured processing of massive data in museums, and algorithms such as ML and RS can reveal the connections between massive artworks, and analyze visual data and text language data of artworks based on ML algorithm, thereby realizing the curatorial work of classifying the nature of the works (Bönisch, 2021). Using an ML algorithm to analyze visual data or language data, determine the data's potential curatorial value, and implement curation is a central theme of these studies. Currently, technology advances faster than its application to curation. As algorithmic technology continues to advance and museum digitization continues to advance, there will be a growing demand for the application of AI algorithms in museum curation. In this context, research in this area has increased over the past few years. In the present study, relevant literatures published between 2004 and 2022 were retrieved from the Web of Science (WOS) database, and then literature metrology and knowledge graph analysis were performed on CiteSpace software to form a knowledge base in the research field to provide the most recent progress, frontier, hot spot, evolution path, and future development trend of museum AI curation research. This research can contribute to the advancement of AI algorithm curation in the museum sector.

METHODS AND MATERIALS

Using bibliometric analysis, mathematics, statistics, and other measurement techniques, the distribution structure, quantity relation, and variation rule of the literature are examined. In addition to articles and books, its analysis objects include other pertinent article information, such as the article's title, subject words, keywords, word frequency, co-citation, co-occurrence, citation information, co-citation references, citation coupling, author, collaborator, publisher, date, language, institution, and country. In this way, it can objectively and exhaustively reveal a field's development and trends, allowing other researchers to comprehend research priorities quickly. In this regard, it can be utilized to analyze a discipline's research survey and development trend.CiteSpace is a Java application for visualizing literature and analyzing co-citations (Chen, 2004). CiteSpace can present abstract data as a graph based on statistical calculations, thereby analyzing indexes of literature co-citation, author co-citation, and keyword co-occurrence (Chen, 2013).

RESEARCH DATA COLLECTION

Using the WOS database as a data source, the retrieval formulas (TS=AI) AND (TS=Algorithm) AND (TS=Museum Exhibitions OR TS=Curation) were used to collect articles published between 2004 and 2022. 577 articles have been screened for the first time, the article type filter is set to journal articles, and the language filter is English. The collected articles with complete records and cited references were then saved as plain text for analysis. It imported the raw data into an Excel spreadsheet and manually verifies each article's publication year, language, title, and type. The committee will reject projects that are not published in English, do not cover the period 2004–2022, or are irrelevant to the topic. Fig. 1 depicts the search strategy flowchart, which produced a data set containing 288 articles. The final data set can be imported into CiteSpace in order to generate the bibliometric analysis graph. Statistical computing illustrates abstract data as a visual map. It consists primarily of literature co-citation analysis, co-citation analysis of journal institution, co-occurrence analysis of subject matter, keyword co-occurrence and keyword cluster analysis, etc.

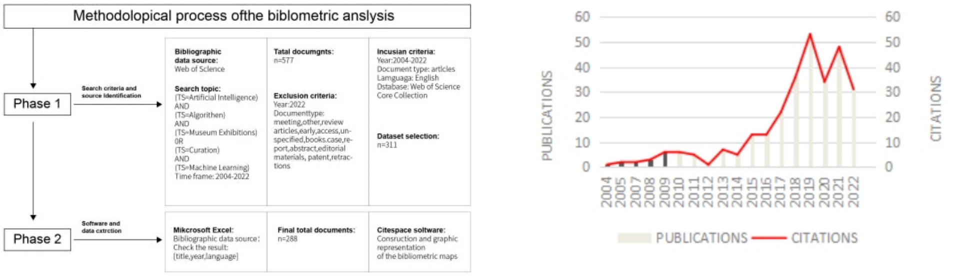

Figure 1: Flow chart of the search strategy and the number of published papers on Museum AI curation (2004–2022).

Parameter Setting

The Set the CiteSpace parameters before processing the data. (1) Select the node type according to the proper analysis; (2) Set the time slice to 2004–2022; (3) Set the length of each time slice to "2"; (4) The selection criterion is TOP N = 50; and (5) pruning was designated as the pathfinder in the selection criteria. The remaining parameters have been assigned their default values.

Analysis of Key Metrics in CiteSpace

CiteSpace employs three metrics (intermediate centrality, contour, and modularity) to assess the Map's logic once it has been generated in accordance with the specified requirements. In terms of structural indicators, betweenness centrality, defined as the ratio of the total number of shortest path lines between two points to the total number of shortest path lines through a point, must be considered (Rousseau et al., 2018). Nodes with a betweenness centrality of greater than 0.1 should be considered critical. When a purple circle appears on the Map, the betweenness centrality is greater than 0.1 (Brandes, 2008). The values of modularity (Q) and average contour (S) are the two most important variables for evaluating the effect of Clustergram. The index of modularity of a network is denoted by the term modularity. Increasing this parameter improves the network clustering results. Q is between 0 and 1. A value greater than 0.3 indicates statistical significance for the split cluster structure. The greater the clustering effect, the closer the value is to 1. The index of Silhouette (S) is utilized to evaluate the network's homogeneity. When it is closer to the value 1, it will more accurately reflect the network's homogeneity. When the value exceeds 0.5, the clustering result can be considered acceptable. Clustering would be effective and convincing in general when the S value is 0.70 (Chen, 2016).

RESULTS AND DISCUSSION

Analysis of Country Cooperation

Select "Country" as the node type in CiteSpace based on betweenness centrality. The 33 nodes and 182 connecting lines depicted in Fig. 2 are obtained. Overall, the country cooperation network in AI curatorial research is closer to the middle countries, and the greater the number of connecting lines, the denser the network. In addition, countries with the highest number of publications are the most cooperative on a national level.

Figure 2: Co-authorship network map of countries publishing on Museum AI curation.

A total of 51 countries participated in the Museum AI curation study between 2004 and 2022. Cluster #0 (Personalized cultural heritage experience—real data) consists of the UK, Greece, and Spain; Cluster #1 (Personalized cultural heritage experience) consists of the US, Italy, and China. Their research in the field of AI curation shares similarities. Each country is represented by a node whose size is proportional to its number of publications. The links between countries are represented by the lines connecting the nodes. The United States and Germany exhibited the largest node diameters, indicating that they are the most influential countries in the field of Museum AI curation. Europe is home to seven of the top ten countries in terms of node diameter, which may be attributed to the region's abundance of cultural heritage museums.

Journal Analysis

Thirty-eight of the 276 journals have been distinguished as meeting the minimum publication threshold of five per journal. Fig. 3 depicts the 38-node periodic citation network. According to Bradford's law, when the number of papers published in a journal's subject area is listed in descending order, the journals in this subject area can be divided into three categories: core field journals, related field journals, and unrelated field journals (Boya, 2016). The formula for calculation is as follows:

$$R_0 = 2\ln\left(e^{E} \times Y\right) \tag{1}$$

Where R_0 is an estimate of the number of journals that should be considered core journals in a given field, and E is the Euler-Mascheroni constant (0.5772) And Y represents the greatest number of articles in the field. In the study, $Y = 73$. Upon calculation, the value of r0 is approximately 9.735 and is rounded to 10. Therefore, at least ten periodicals should be considered as core periodicals in the field of Museum AI curation. There are ten leading research journals in the field, including Lecture Notes in Computer Science and Information Communication & Society. These journals originate from the first three clusters depicted in Fig. 6's network cluster, namely cluster #0 (social media), cluster 1 (museum visitor), and cluster #3 (ontology technique).

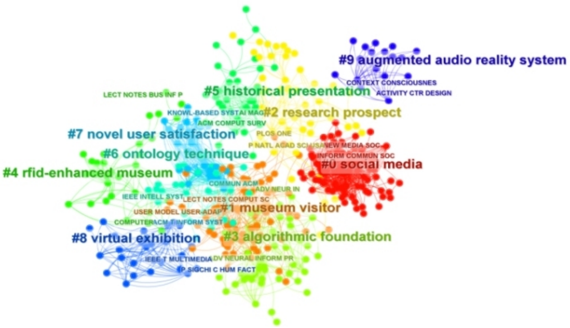

Figure 3: Network visualization map of citation analysis of journals publishing on AI curation.

Cluster #1's Lecture notes in computer science (2004) have the most citations, as shown by Table1. The number cited is 73. Cluster #0's second entry is Information, communication, and Society (2016). The number cited is 48. Communications of the ACM (2004) is the third item in cluster #3. The number cited is 42. The fourth is New Media & Society (2017) in cluster #0, where the number cited is 37. Fifth in cluster #1 is User modeling and user-adapted interaction (2005). The number cited is 33. Cluster #0's sixth item is Digital journalism (2016). The number cited is 32. Information Communication & Society, New Media & Society, and Digital Journalism all belong to Cluster #0, indicating that "social media" is the focus of their research. Cluster #1 consists of Lecture notes in computer science and User modeling and user-adapted interaction, which focus on the related research of "museum visitor"; Cluster #3 consists of Communications of the ACM, which focuses more on the study of ontology technique in Museum AI curation.

CiteSpace's analysis revealed that research on AI curation in museums spans multiple disciplines, including information science, museology, library science, computer science, art and humanities, urban studies, anthropology, etc. Social Media, Museum Visitor, and Ontology Technique have the most literature among them. This demonstrates that contemporary museum curation and museum AI curation are multidisciplinary fields of study. Computer science lecture notes ranked first in terms of citation frequency. It was discovered that the curation of the content of the museum's digital collections is heavily influenced by the interaction between digital museum visitors and computers The second and third-ranked journals both originate from Cluster #0 (social media). Social media platforms are increasingly characterized by algorithm-driven content management. Consequently, public curation of AI technology in museum social media is currently one of the most important AI curation research directions.

Table 1. Journals in the core area of AI curation (ranking by articles).

Journals	No.	Year	Clusters ID	Silhouette	Clustering Labels LLR
Lecture notes in computer science	73	2004	#1	0.702	museum visitor
Information communication & society	48	2016	#0	0.902	social media
Communications of the ACM	42	2004	#3	0.925	ontology technique
New media & society	37	2017	#0	0.902	social media
User modeling and user-adapted interaction	33	2005	#1	0.702	museum visitor
Digital journalism	32	2016	#0	0.902	social media
Computers in human behavior	30	2013	#0	0.902	social media

Keyword Co-Occurrence and Keyword Cluster Analysis

Analysis of frequently cited literature Keywords represent the core content of the literature, and keywords with a high frequency reflect the research hotspot in this field. Each keyword is represented by a node whose size is proportional to its frequency, More links indicate a greater frequency of keyword co-occurrence. The thickness of the connection indicates the strength of the connection. The total network consists of 317 nodes and 862 links. CiteSpace has extracted cluster labels of research terms at various locations in cited documents using the Log-likelihood rate (LLR) and Mutual information (MI) algorithms. The evolution of research hotspots in this field between 2004 and 2022 is depicted in Fig. 4. It depicts the time frame and research process of the development and evolution of each research keyword clustering. It focuses primarily on describing the connection between clusters and the time span of literature within a cluster. Each cluster within the graph represents a time axis. Each node on the axis represents a keyword, with the size of the node representing the keyword's occurrence frequency. The position of each node indicates the occurrence of the corresponding keyword. In this analysis, a total of ten keyword clustering labels with the cluster number and identifier were generated. The larger the LLR contour value for keyword clustering of 10 clusters, the more representative the word is for this cluster. Each cluster or cluster combination is representative of a museum AI curation subfield.

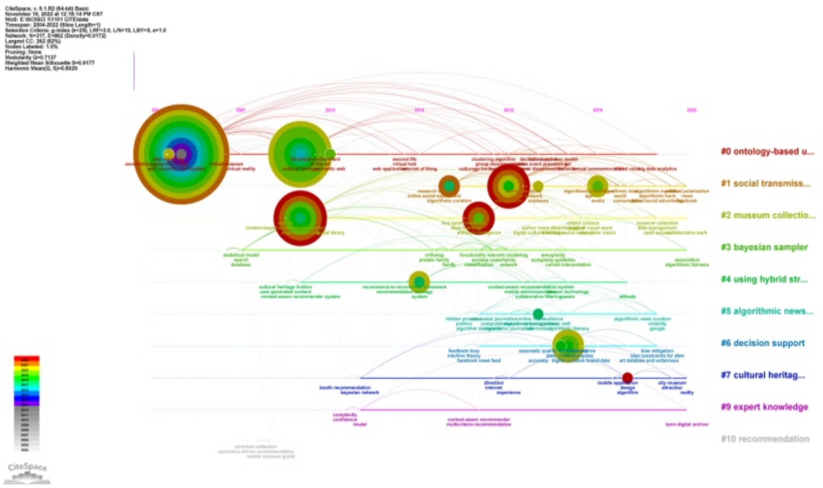

Figure 4: Domains of human systems integration (adapted from U.S Air Force, 2005).

The top ten most-used keywords in this field can be sorted using. Specifically, recommender system (45) has the highest frequency, followed by ML (27), indicating that RS is the most important research object in this field. The top 15 Digital humanity-related keywords are "cultural heritage," "social media," "AI," and "system." Digital humanity adopts the methodology of the social sciences and computing tools, such as hypertext, data visualization, statistics, text exploration, digital mapping, etc. This indicates that digital humanity is one of the most popular research techniques in this field. Other frequently occurring keywords are "algorithmic curation", "digital curation", "augmented reality", "big data", "data sharing", "model", and "natural language processing (NLP)". These keywords revealed that algorithms such as ML and NLP in the field of AI have provided museum AI curation with specific technologies, visual data and text language data analysis, keyword semantic analysis, and other methods. Big data has produced meta-data and canonical training sets that can be applied to this field's research. “Social media” and “cultural heritage” provide meaningful content and entry points for intelligent curation research and development.

Analysis of Frequently Cited Literature

Ranking articles by citation is a traditional bibliometric technique that can reveal the most influential articles in a field. The most-cited article is the emotional dimension and perceptual experience of algorithms by analyzing the kinds of scenes and spaces in which people and algorithms interact, focusing on personal stories generated by algorithms on social media. It should be noted that algorithms can generate stories and emotional experiences, and that user engagement has a formative effect on social media algorithms (Bucher, 2017). Ten times, the article “Thinking critically about and researching algorithms” was cited. First, algorithms are analyzed critically and empirically, and the optimal method for studying them in practice is discussed. There are numerous ways to conceptualize algorithms (technology, computing, mathematics, politics, culture, economics, context, matter, philosophy, and ethics). Nonetheless, it should be viewed as accidental, ontogenetic, and performative, and as part of a larger sociotechnical portfolio. In addition, six methodological approaches designed to provide insight into the nature and operation of algorithms are evaluated critically (Kitchin et al., 2017). Although these two articles are not primarily concerned with the museum’s AI curation as the research object, they do discuss the nature of algorithms and the rise of “algorithmic machines” New forms of algorithmic capabilities are reshaping the way in which social and economic systems function and conducting theoretical and ethical research. Awareness of algorithms has the potential to be a critical issue (Gran et al., 2021). The power, opacity, and bias of algorithmic systems have opened up new research areas for bringing transparency, fairness, and accountability into these systems (Eslami, 2021). As a result, it has become one of the most frequently cited articles in the field.

DISCUSSION

On the basis of the mapping of network data of AI curation literature by the CiteSpsce software, the relationship between authors, countries, and institutions in this field, as well as the co-occurrence and co-citation of keywords, research hotspots, and trends, are determined in the present study. We provided a summary of research on AI curation: It is based on ontology–based user modeling and “Bayesian sampler”, “using hybrid strategies”, “algorithmic selection” and “recommendation” set up by the technology of AI technology and the method dimension; for another, it is the curatorial content dimension based on “museum collection” and “expert knowledge” and “social transmission, decision support, cultural heritage” The public aspect portrayed by keywords such as “visit.” From these three dimensions, AI curation and curation research have been materialized. In addition, AI algorithm curation has not yet been studied systematically, especially in terms of the public’s use of AI technology to intervene in museum cultural heritage exhibition content curation research. Moreover, algorithm is detrimental. This is another issue that must be addressed and resolved during the algorithm curation process. As the algorithm is opaque to the user, technology improves the manner in which humans create knowledge in the digital age and influences a more "real" experience for users. Audiences believe that algorithmic selection based on past user behavior is a more efficient way to acquire knowledge than editorial curation. Nonetheless, these beliefs vary significantly on an individual level. It displays background characteristics that have sociological implications for algorithmic understanding.

CONCLUSION

Intelligent curation research focuses on revealing the network of interactive digital museum and library collections, databases, expert knowledge and ML, and recommendation systems. In addition, it also includes your research on critically evaluating the power and social dynamics embedded in data within AI systems. There are currently a number of research articles on AI curation in the museum field. However, the current most-cited literature in this field focuses primarily on the technology, methodology, and ethics of AI curation, whereas there are few publications on the specific methods and pathways of AI curation in museums. Research on the technical methodology and ethics of algorithmic curation is one potential direction of development in this field; the practice of intelligent curation can be guided by the iterative development of new algorithm models. Future studies can concentrate on content innovation. Moreover, with the application of new technology and data analysis techniques, the ethical research of algorithm curation from a sociological and anthropological perspective is also a popular research direction among academics. curation functions within a wider ecology of social and technical power relations. Overall, a well-designed and formidable system can facilitate collection exploration, and it can also be used to retrieve data that cannot be located using conventional methods or that is not necessarily interrelated in a traditional curation (thematic or chronological aspects). For instance, we have proposed a virtual exhibition authoring tool that is capable of guiding users from querying knowledge graphs to automatically generating virtual experiences, providing comprehensive support for users throughout the entire creative process.

ACKNOWLEDGMENT

This work was supported by the 2022 Guangzhou Philosophy and Social Science Development Project - Research on Intelligent Generation of Digital Museum Exhibition in Guangdong-Hong Kong-Macao Greater Bay Area from the Perspective of Philosophy of Technology (2021GZGJ294) and 2021 Guangdong Degree and Postgraduate Education Reform Research Project - Research on Cultivation Model of Design Research Ability in the Integration of Art and Science (2021JGXM077).

REFERENCES

Anne-Britt Gran, Peter Booth & Taina Bucher. (2021). To be or not to be algorithm aware: A question of a new digital divide?, Information, Communication & Society, 24:12, 1779–1796.

B. A. Campbell. (2019). "Museum collections: An overview," The curation and care of museum collections pp. 1–24.

C. Chen. (2004). "Searching for intellectual turning points: Progressive knowledge domain visualization," Proceedings of the National Academy of Sciences, vol. 101, no. suppl_1, pp. 5303–5310.

C. Chen. (2013). "The Structure and Dynamics of Scientific Knowledge," in Mapping Scientific Frontiers: The Quest for Knowledge Visualization, C. Chen Ed. London: Springer London, pp. 163–199.

C. C. Chen. (2016). A Practical Guide for Mapping Scientific Literature. Hauppauge, NY, USA: Nova Science Publishers.

D. Bönisch. (2021). "The Curator's Machine: Clustering of Museum Collection Data through Annotation of Hidden Connection Patterns between Artworks," International Journal for Digital Art History, no. 5, pp. 5.20–5.35.

F. R. Cameron. (2021). The Future of Digital Data, Heritage and Curation: In a More-than-Human World (1st ed.).

K. Thorson and C. Wells. (2016). "Curated flows: A framework for mapping media exposure in the digital age," Communication Theory, vol. 26, no. 3, pp. 309–328.

L. E. R. Rousseau, and R. Guns. (2018). Becoming Metric-Wise: A Bibliometric Guide for Researchers. Chaandos Publishing, Oxford, UK.

M. Tyzlik-Carver. (2016). "Curating in/as commons posthuman curating and computational cultures," Doctor thesis, Aarhus University.

M. Eslami et al. (2016). "First I "like" it, then I hide it: Folk Theories of Social Feeds," in Proceedings of the 2016 cHI conference on human factors in computing systems, 2016, pp. 2371–2382.

M. A. DeVito, D. Gergle, and J. Birnholtz. (2017). "Algorithms ruin everything # RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media," in Proceedings of the 2017 CHI conference on human factors in computing systems, pp. 3163–3174.

Motahhare Eslami. (2021). Revisiting Transparency and Fairness in Algorithmic Systems Through the Lens of Public Education and Engagement. In Proceedings of the Eighth ACM Conference on Learning @ Scale (L@S '21). Association for Computing Machinery, New York, NY, USA, 13.

N. Diakopoulos. (2019). Automating the News: How Algorithms Are Rewriting the Media. Cambridge: Harvard University Press.

N. Thurman, J. Moeller, N. Helberger, and D. Trilling. (2019). "My friends, editors, algorithms, and I: Examining audience attitudes to news selection," Digital journalism, vol. 7, no. 4, pp. 447–469, 2019.

R. Kitchin. (2017) "Thinking critically about and researching algorithms," Information, communication & society, vol. 20, no. 1, pp. 14–29.

U. Brandes. (2001). "A faster algorithm for betweenness centrality," Journal of Mathematical Sociology, vol. 25, no. 2, pp. 163–177.

W. Z. L. Boya. (2016). "Dynamic analysis of information science research focus of the past five years in China: Based on Bradford law partition theory," Information and Documentation Services, vol. 37, no. 3, pp. 34–40.