Exploring different notions of literacy: a literature review analysis of literacy research related to Artificial Intelligence and Big Data application

Zijia Wang*  
1New Media Research Center, Sun Yat-sen University Nanfang College, Guangzhou, Guangdong, 510970, China  
*Corresponding author’s e-mail: wang.zijia@outlook.com

Abstract. Artificial Intelligence (AI) and Big Data projects have been rapidly applied by organizations and adopted by younger users, leading to an increasing number of various capacity building applications, as well as significant impacts on the well-being of both designers and users. This research presents a review of literacy-related literature on “Big Data”, “Artificial Intelligence”, “Machine Learning”, based on 73 articles retrieved from the Web of Science Database. The author produced a keyword co-occurrence network map highlighting three main domains: digital literacy, information literacy, and data literacy. This paper concluded that data and information literacy share similar semantic networks whereas digital literacy is related more to pedagogy and innovations, while being the more distant node from both data and information literacy. For stakeholders such as educators, academic researchers, and policymakers, the paper provides reference points for understanding technological literacy and the current relationship among key literacy concepts and domains in relation to AI and Big Data.

1. Introduction  
The industrial applications, our economies, and even education are rapidly transformed by rapid technological advancement in big data, AI and machine learning [1]. Such domains of technological development also make great potential to fight hunger, climate crisis, etc. towards a more sustainable future [2], which brings great value to educate younger generations and overcome the digital divide [3]. In the previous literature, Nuortimo (2018) has conducted a review on digital transformation in the Mining industry [4]; MacMillan (2014) has reviewed data management and the sharing of librarians [5]. The existing literature has yet covered the impact of emerging technology trends such as big data, AI or machine learning towards literacy. This paper aims to provide a literature review on the specific topic of literacy and its relations with AI and Big Data applications.

To examine the relationships between the notion of literacy, a key concept for education and educators, and Big Data and AI applications, this paper conducted a literature review of 72 retrieved articles from the Web of Science (WoS) database. The purpose of this article is to explore the potential links and gaps among different domains in literacy and technologies for educators, researchers and policymakers.

2. Data and methods  
The paper first describes the query design that defines literacy in relation to big data, AI and machine learning, and then explains the mapping protocols and strategies as follows.
First, based on the entries of “digital literacy” and “literacy” from Oxford Bibliographies [6,7], a series of related terms have been compiled, ranging from data literacy, digital literacy to multiliteracies.

Then from the WoS database, we executed queries below to gather literature data on the various terms of literacy in relation to Big Data and AI on Nov 2019:

- \(\text{TS = ("Artificial Intelligence" OR "machine learning" OR "Big Data") AND TS = ("data literacy" OR "digital literacy" OR "information literacy" OR "internet literacy" OR "computer literacy" OR "ICT literacy" OR "Information Communication and Technology literacy" OR "Information Communication & Technology literacy" OR "media literacy" OR "new literacy" OR "multiliteracy" OR "multiliteracies")}\)

As a result, 73 articles are retrieved in November 2019. VOSviewer and Python visualization packages was used for mapping.

3. Research mapping results

3.1. Annual trends

Figure 1 shows that the number of publications has been scarcely from 2000 toward 2013. However, the number starts to increase rapidly since 2013 and keeps on growing steadily. Figure 2 shows the annual citation numbers, revealing the heightened interests and perhaps paper quality on the topic.

![Figure 1. Publications Trend from 2000 to 2019](image1)

![Figure 2. Citations Trend from 2001 to 2019.](image2)

3.2. Top WoS categories, research areas and organizations

The top WoS categories are listed in Table 1, with Information Science & Library Science, Education & Educational Research, and Computer Science, Information Systems as the top three. The top three categories make up about one-third of the total publications, indicating clustering in the fields of information, library and education related research. Some other categories include social sciences and interdisciplinary applications.

| WoS categories                          | No. of Articles | 136 in total |
|----------------------------------------|----------------|--------------|
| Information Science & Library Science  | 21             | 15%          |
| Education & Educational Research       | 15             | 11%          |
| Computer Science, Information Systems  | 11             | 8%           |
| Computer Science, Theory & Methods     | 7              | 5%           |
| Social Sciences, Interdisciplinary      | 5              | 4%           |
| Computer Science, Artificial Intelligence | 5            | 4%           |
| Computer Science, Interdisciplinary Applications | 5     | 4%           |
| Computer Science, Software Engineering  | 5              | 4%           |
| Engineering, Electrical & Electronic   | 4              | 3%           |
| Education, Scientific Disciplines      | 4              | 3%           |
The top research areas are shown in Table 2. Of the 13 areas that have more than 2 articles, the top three are the same as WoS categories. Those three same areas make up 52% of total publications, which indicate a more concentrated clustering on computer and information sciences. More diverse research areas have appeared in WoS categories such as business, government & law, public administration, etc.

Table 2. Top research areas

| Research areas                          | No. of Articles | %  |
|----------------------------------------|-----------------|----|
| Computer Science                       | 23              | 19%|
| Information Science & Library Science  | 21              | 18%|
| Education & Educational Research       | 18              | 15%|
| Engineering                            | 9               | 8% |
| Business & Economics                   | 5               | 4% |
| Social Sciences - Other Topics         | 5               | 4% |
| Telecommunications                     | 3               | 3% |
| Government & Law                      | 2               | 2% |
| Physical Geography                     | 2               | 2% |
| Public Administration                  | 2               | 2% |
| Remote Sensing                         | 2               | 2% |
| Operations Research & Management Science| 2               | 2% |
| Communication                          | 2               | 2% |
| Others                                 | 24              | 20%|

Table 3. Top Countries

| Countries     | No. of Articles | %  |
|----------------|-----------------|----|
| USA            | 18              | 23%|
| China          | 15              | 19%|
| England        | 10              | 13%|
| Germany        | 5               | 6% |
| Canada         | 4               | 5% |
| Indonesia      | 3               | 4% |
| Greece         | 2               | 3% |
| Hungary        | 2               | 3% |
| Netherlands    | 2               | 3% |
| Scotland       | 2               | 3% |
| Spain          | 2               | 3% |
| Others         | 14              | 18%|

3.3. Top countries

Table 3 lists all countries that have two articles. The top three regions are USA, China and England, which makes up 55% in total. Among them, Indonesia stands out as the main developing country other than China.

3.4. Top publications

Among the top-cited articles, Hargittai argued that the biased sampling frames of social network analysis could lead to methodological challenges of big data research [8]. The other top 10 most cited articles mainly focused on the impact of data/information literacy on various domains such as accounting [9], undergraduate chemistry curriculum [10], nurse shift workers [11], and librarians [12], etc. No clear author clusters were found, implying a very decentralized citation networks without top or star scholars.

3.5. Keyword map and Clusters

This research analysed occurrences and produced a co-occurrence network based on the keyword data, resulting in a concise keyword map due to the relatively small number of total articles from the database. VOSviewer was used to visualize the network, in order to detect the relationships among keywords. Figure 3 shows the overall keyword map based on co-occurrence data. The keywords analysis network is consisted of five clusters: (1) machine learning and digital literacy and, (2) artificial intelligence, data and librarians, (3) big data and data literacy, (4) information literacy and education, and (5) technology. Technology cluster contains only one node at the center of the network. Three types of literacy belongs to different clusters, echoing the complexity of defining terms such as “data literacy”[13]. Among the three types of literacy, digital literacy appears to be peripheral (i.e. further away from the more closely situated data literacy and information literacy). Keywords such as “innovation” and “technology” serve as nodes so that connect digital literacy to the core network.
Figure 3. overall keyword map based on co-occurrence data

As shown in (a), (b), (c) of Figure 4, three technology-related keywords stand out on the map. Among them, “big data” shares the most intensive connections compared with “artificial intelligence” and “machine learning”. Three literacy-related keywords stand out as main connections among these technologies, which are “information literacy”, “digital literacy” and “data literacy” shown in (e), (f), (g) of Figure 4. Big data shares a connection with all three domains of literacy; machine learning only relates to information and digital literacy, yet AI relates to none. Both AI and machine learning are linked to librarians and pedagogy.

4. CONCLUSION
The paper intends to contribute to the body of knowledge towards the relationship between domains of literacy and technological keywords. With the aim to understand the links and gaps of literature related
to literacy, big data, AI and machine learning, this paper visualizes the keyword analysis networks of key publications from the WoS database. In relation to literacy, big data and machine learning are in more close links compared with AI, implying great potential in the promotion of AI literacy. In addition, data and information literacy share part of networks, indicating some similarity and overlapping use of both concepts. While information literacy relates to knowledge and technology, and data literacy relates to smart city and pedagogy. As a relatively independent cluster from data and information literacy, digital literacy shares more connections with keywords such as management, innovation and e-learning, revealing more separate domains with potential connections with business and online education.

The results of this paper have revealed the gaps and potentials to better understand the dynamics of different literacy with technologies, providing reference points for stakeholders such as educators, academic researchers, and policymakers who are interested in technological literacy and the connections among the key literacy domains in relation to AI and Big Data application.

**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 HCI for Education Development (2018WQNCX283), under the 2018 Major Projects of Guangdong Province Department of Education "Young Innovative Talents" Grants.

**References**

[1] UNESCO (2019) Principles for AI: Towards a Humanistic Approach? UNESCO (Paris: UNESCO)

[2] UN News (2019) Artificial intelligence summit focuses on fighting hunger, climate crisis and transition to 'smart sustainable cities.' UN News

[3] Schweitzer, E.J. (2019) Digital divide. SOCIETY

[4] Nuortimo, K. (2018) Measuring public acceptance with opinion mining: The case of the energy industry with long-term coal R&D investment projects. J Intell Stud Bus 8.0: 6.

[5] MacMillan, D. (2014) Data Sharing and Discovery: What Librarians Need to Know. J. Acad. Librariansh. 40: 541–9.

[6] Hicks, T., Baleja, K., Zhang, M. (2019) Digital Literacy. Communication

[7] Rennie, J., Ortlieb, E. (2013) Literacy. Education 9780199756810–0095.

[8] Hargittai, E. (2015) Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites. Ann. Am. Acad. Pol. Soc. Sci. 659: 63–76.

[9] Bhimani, A., Willcocks, L. (2014) Digitisation, ‘Big Data’ and the transformation of accounting information. Account. Bus. Res. 44: 469–90.

[10] Reisner, B.A., Vaughan, K.T.L., Shorish, Y.L. (2014) Making Data Management Accessible in the Undergraduate Chemistry Curriculum. J. Chem. Educ. 91: 1943–6.

[11] Hobbs, B.B., Farr, L.A. (2004) Assessing Internet Survey Data Collection Methods with Ethnic Nurse Shift Workers. Chronobiol. Int. 21: 1003–13.

[12] Koltay, T. (2016) Data governance, data literacy and the management of data quality. IFLA J. 42: 303–12.

[13] Koltay, T. (2015) Data literacy: in search of a name and identity. J. Doc. 71: 401–15.