

# Analysis on the Model of Data Literacy Cultivation in Universities in the Age of Big Data

Jun Li<sup>1</sup>, Xiangping Zhang<sup>1</sup>, Shengmei Zhang<sup>2,\*</sup>

<sup>1</sup>The School of Language and Media, Anhui University of Finance & Economics, Anhui, Bengbu, 233030, China

<sup>2</sup>The School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, 233030, Anhui, China

\*smzhang001@126.com

## Abstract

With the rapid development of data technology, the Big Data has become an indispensable part of our work and life. And the corresponding data literacy has also become one of the basic qualities that citizens should have in the future. As an important force in the future social development, college students should have higher data quality. At the moment, it is not clear on the methods and ways of College Students' data literacy education. Many subjects and primary and secondary relations are the main obstacles that restrict the training of data literacy of college students. So to explore a path for training college students on data literacy, we should think from the big data's production and demand. University library, professional teaching department, campus information institutions and students should form a four-in-one education method.

## Keywords

Data literacy; Four-in-one; Cultivation model.

## 1. Introduction

In 2012, Gray Brink [1] declared in a column in the New York Times that the "age of big data" had arrived. Since then, big data has rapidly become a global topic. The era of big data has quietly arrived. However, in China, the overall data literacy of ordinary citizens is still at a relatively low level. As the important inheritor of the baton of social development, the current situation of data literacy of college students is not optimistic. In this paper, we start from the connotation and composition of data literacy and try to explore the way of constructing data literacy education model in the era of big data in colleges and universities.

## 2. Definition of the Concept and Content of Data Literacy

### 2.1. Evolution of the Concept of Data Literacy

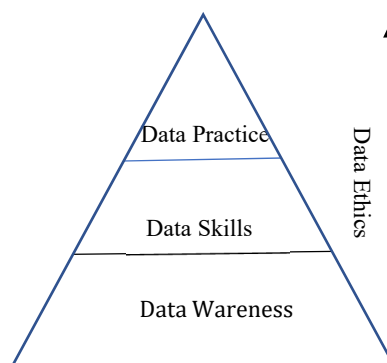
The concept of data literacy first originated in the educational community in the United States. In 2007, Stevenson and Caravello [2] defined data literacy as the ability to find, evaluate, and use information (including data resources) effectively and rationally. In 2011, Hogenboon [2] and others proposed that data literacy is the ability to read, interpret, analyze, and think critically about statistical data and the ability to use statistical data as evidence. Domestic scholars have also provided their insights on data literacy. According to Yuan-lingHao and Ting-ting Shen [3], data literacy is the ability to have data awareness and data sensitivity, to acquire, analyze, process, use and present data effectively and appropriately, and to think critically

about data, which is an extension and expansion of statistical literacy and information literacy. According to Qing Li and Shu Ren [4], data literacy is the ability of people to manage, interpret, and think critically about the ever-growing amount of data.

From the research results that have been achieved on data literacy in the academic community, the conditions for a complete and standardized conceptual definition of data literacy are not yet available. This paper also does not intend to propose a new concept. However, based on the current situation of the use of big data scale and technology development, we can make some generalizations about the content of data literacy as follows.

## 2.2. Content of Data Literacy

On the basis of the consensus that big data development has become a national development strategy level, different social groups have different requirements for the data literacy they should have. However, in general, data literacy should cover the following aspects (as shown in Figure 1).



**Figure 1.** Basic components of data literacy.

### 2.2.1. Data Awareness

Data awareness includes data sensitivity, concern, critical thinking about data itself, and the ability to judge the comprehensive value of data, data security, and the significance of sharing. Without data sensitivity and attention, individuals may quickly face a widening data knowledge gap, which ultimately affects personal development.

Among the components of data literacy, data awareness is a prerequisite for the formation of data literacy, as data are widely distributed in life, large in scale and diverse in form. Different data streams have their own trajectory. If the existence and value of data streams are not fully recognized, there is no way to collect and organize data. It is also impossible for big data to play its proper role.

While recognizing the important utility of data, we should also maintain critical thinking about data. Data is not omnipotent; it is only a tool that reflects the objective world and helps people discover the laws of its operation. Data omnipotence, or even data worship, can only be the paranoia of data fetishists.

### 2.2.2. Data Ethics

According to Kranzberg, "Technology is neither good, nor bad, nor neutral ..... The impact of technological developments on the social ecology often goes beyond the technological devices and practices themselves." [1] Similarly, while enjoying the convenience of big data, one should be vigilant about the ethical and legal risks that may arise from big data technologies.

Data ethics should include. First, it is necessary to use big data acquisition technology reasonably to avoid infringing on the privacy and legitimate interests of data owners (individuals, social groups, commercial organizations, etc.). Second, work with data within the permissible framework of laws and regulations. Third, the use of data must not interfere with civil liberties and social justice.

Big data should have sufficient social publicness, and the content of big data should eventually become a public product for the whole society. Privatization of big data, or exclusive enjoyment by certain organizations, will eventually result in the continuous loss of the public's right to speak about their social public activities. Therefore, data ethics should run through the whole process of data discovery and collection, processing, communication and publication, and should become the basic content of data literacy.

### **2.2.3. Data Skills**

Data skills include the technical ability to collect, acquire, select, process, store, communicate, and publish initial data. With the rapid development of new technologies, data is increasingly presented in various forms. It can not only be visualized and displayed as images, but can even be transformed into people's favorite art style. Behind all these is the improvement of people's comprehensive data skills. With the continuous development of technologies such as automated data capture, cloud platform storage and exchange, data analysis and data visualization, data has become more and more closely integrated with real life. Whether in education and teaching, or in medical technology, financial innovation, or IoT development, the comprehensive data processing skills have been reflected accordingly.

### **2.2.4. Ability to Handle Data in Practice**

The rational use of big data also requires the ability to closely integrate data with real-life practical problems, including: the ability to use data to understand objective things; the use of big data technology to clearly explore the development pattern of social affairs and natural conditions. The ability to recognize the development trend of things through appearances; the ability to identify problems and provide insights for people to solve problems based on the former; the ability to reuse and recreate data. As a kind of metadata, the big data created directly by the public has the ability to reproduce social life in all aspects, and its value cannot be fully explored by one-time use. We should pay attention to the reuse and re-creation of data.

## **3. The Current Situation and Problems of Data Literacy Cultivation in Universities**

### **3.1. Initial Establishment of Talent Cultivation System**

In 2015, the State Council put forward the "Action Plan for Promoting the Development of Big Data", pointing out that data has become a basic strategic resource of the country, and big data is increasingly having an important impact on global production, circulation, distribution and consumption activities, as well as economic operation mechanism, social life style and national governance ability. For the training of big data talents, the State Council proposed to innovate the talent training model, establish and improve the multi-level, multi-type big data talent training system. Encourage universities to set up data science and data engineering related majors.

In February 2016, the Ministry of Education added the professional direction of "data science and big data technology", and by June 2017, 35 universities in China have established big data and related majors. Some other universities have added data technology and its related majors in addition to the original professional catalog. In the higher level of postgraduate training, some universities have also established exclusive data science institutes or big data colleges, or added new related directions under the original master's specialties. For example, Guizhou University has set up a big data institute, Tsinghua University has set up a data science research institute, and Peking University has jointly established a collaborative innovation platform for the cultivation of master's degree in big data analysis with Renmin University of China, University of Chinese Academy of Sciences, Capital University of Economics and Business, and

Central University of Finance and Economics, which has been enrolled since September 2014. Fudan University, Beijing University of Aeronautics and Astronautics, Shanghai University of Finance and Economics and some other universities have also enrolled undergraduate or master's students related to big data in September 2014 and beyond. It can be said that after the discipline development in recent years, China has initially established a comprehensive training system for data technology and big data majors. The cultivation of data literacy in colleges and universities has also entered a new stage.

### **3.2. Problems Facing the Cultivation of Data Literacy Talents**

Although the cultivation system of data literacy in colleges and universities has been initially built, the existing cultivation method of data literacy is still dominated by the traditional teaching cultivation method. This cultivation method has shown some shortcomings in the present time when data technology is fully penetrated.

#### **3.2.1. Insufficient Diversification of Teaching Forms**

Looking at the data literacy courses offered by colleges and universities, we can find that the construction of the existing data literacy course system is not sufficient, and the teaching form is mainly based on traditional class teaching, supplemented by certain online teaching forms such as mulberry courses. However, in the online course teaching, it is still dominated by the traditional teacher's simple lecture, and the teacher-student interaction is still relatively low [5].

Therefore, in the future teaching of data literacy, the teaching body should continuously improve the teaching format according to the development of big data technology and Internet technology. We should make full use of multiple media, promote the deep integration of teaching channels such as classroom, Internet and mobile media, and enhance the teaching effect with more innovative technical means.

#### **3.2.2. Insufficient Updating of Teaching Concept and Teaching Content**

In the comprehensive cultivation of data literacy of college students, the existing teaching mode still takes colleges and universities as the main body of data talent cultivation. The contents taught in teaching are mostly focused on data skills, but less on data ethics and data practice. The fundamental problem of this situation lies in the insufficient updating of the education concept based on data literacy [6]. For example, the training program of the 2018 class of data science and big data technology in a domestic university reflects the situation that more courses are set for data skills and fewer courses are arranged for data ethics and data practice [7].

#### **3.2.3. The Teaching Object Is Relatively Single**

Under the existing teaching mode, the teaching targets are mostly students majoring in data technology and big data, and insufficient attention is paid to the improvement and cultivation of data literacy of college students in other majors.

In 2014, the United Nations proposed in the declaration of "The road to dignity by 2030: ending poverty, transforming all lives and protecting the planet" (The road to dignity by 2030: ending poverty, transforming all lives and protecting the planet) The world needs a new ability - data literacy. As the main force and important builder of future social development, university students should also have the corresponding data literacy.

#### **3.2.4. The Teaching Subjects Are Not Clear and Not Interactive Enough**

In the cultivation of college students' data literacy, there are multiple subjects. However, in the specific cultivation process, the existing cultivation mode is still mostly based on the cultivation of professional departments in universities, and the role of other subjects is not sufficiently recognized. The multiple subjects of data literacy cultivation tend to participate in the teaching of data literacy of college students spontaneously only from their own data needs and the needs

of data talents. There is a lack of mutual cooperation and mutual integration among the subjects, and the stable interactive relationship among the subjects has not been fully established. Specifically, the cross-regional and interdisciplinary comprehensive cultivation between universities and colleges is weak, the quantity and strength of joint cultivation between social information institutions and universities are slightly weak, and the interactive relationship between students and social information institutions is at a low level. All these need to be solved by the continuous integration and strengthening interaction among the subjects.

#### 4. Multi-subject Collaboration: The Cultivation Mode of Data Literacy in Colleges and Universities

In the process of data literacy cultivation of college students, the determination of teaching subjects and the confirmation of mutual relationship should be the premise of carrying out integration and interaction, mutual cooperation, so as to improve the cultivation effect. The determination of teaching subjects should be considered from the dimensions of data generation, data technology and data talents' demand. In terms of the existing cultivation mechanism for students in our universities, the cultivation of data literacy should include four subjects. They are: university libraries, professional teaching faculties, off-campus information institutions and school students. Libraries, teaching faculties and off-campus institutions are the main subjects of education practice, and school students are the main subjects of education (as shown in Figure 2). Only with the joint efforts of the four parties can we effectively improve the cultivation level of data literacy in colleges and universities.

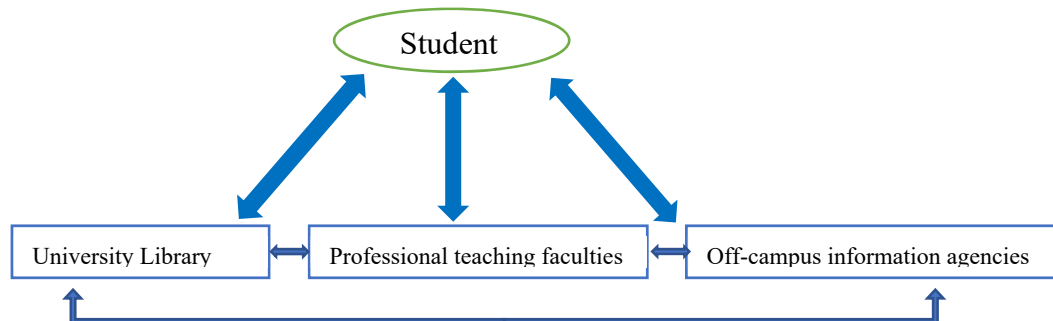


Figure 2. Schematic representation of the four-in-one data literacy development model

##### 4.1. University Libraries

Libraries assume the comprehensive functions of data storage, data sharing and usage in colleges and universities. As the hub of data acquisition, storage, exchange and use, university libraries naturally become an important subject of data literacy cultivation in universities.

In fact, since the Ministry of Education issued the Notice on the Opinions of Opening "Literature Retrieval and Utilization" Class in Higher Education Institutions in February 1984, libraries have been playing a significant role in the data literacy education for students in domestic universities and have been in the forefront of research on information literacy and data application. Libraries have also made considerable contributions to the development of data literacy for college students. Some of the teaching approaches and methods adopted are: online teaching, offline teaching, symposiums, expertise consultation and provision of practical places and laboratory resources [8]. It should be said that university libraries have in fact undertaken the main work of data literacy education in universities before the large scale introduction of big data related courses in professional teaching faculties.



## 4.2. Professional Teaching Faculties

By 2017, 35 universities in China have set up "data science and big data technology" or related majors, and the professional direction involves big data development, data mining, data analysis, machine learning, data operation and maintenance, and cloud computing. Since then, the cultivation of data literacy in colleges and universities has entered a new brand stage. The teaching faculty will gradually become the main force of data literacy cultivation for college students.

However, because the content of big data involves many professional fields and the objects are wide and diverse, other teaching faculties are still the important force of data literacy education in colleges and universities, in addition to the cultivation of professional big data-related faculties. Other teaching departments should improve their comprehensive understanding of big data and fully expand the professional field of data literacy education on the basis of in-depth cultivation in professional fields.

In general, the teaching faculties of universities belong to different professional fields, and they have many specialties in their respective professional fields and have very strong strength in the professional application field of big data. Moreover, teaching faculties not only study big data in depth in scientific research, but also have urgent requirements for data acquisition, storage and use. This has become an intrinsic motivation for teaching faculties to actively participate in data literacy training in universities.

## 4.3. Off-campus Information Agencies

Off-campus information agencies should include two types of subjects: One is government departments. Government departments are responsible for social management and social services, and have priority access to data, and they have acquired a large number of data samples and rich data resources in their daily work. These precipitated data need both relevant professional background and high data literacy staff to deal with the maintenance. However, the fact is that there is a shortage of more than 1.5 million professionals in the field of big data in China [9]. A large portion of the massive data held by government departments still cannot play its proper role. Because of this, Xi Jinping emphasized on December 8, 2017 [10] that big data should be used to enhance the modernization of national governance, establish and improve mechanisms for big data-assisted scientific decision-making and social governance, promote innovation in government management and social governance models, and achieve scientific government decision-making, accurate social governance, and efficient public services.

Second, social information agencies. Social information service institutions and some commercial companies, information companies, technology companies, etc. have data resources of different scales in the era of big data. These data resources have different objects, different scales, and different ways of handling them. They are both the possessor and the consumer of data resources. It becomes an inevitable choice for social organizations to use data resources and universities to cooperate and conduct school-enterprise integration and cultivation to expand the competitiveness of their own data and cultivate new profit growth points.

## 4.4. Students in Colleges and Universities

The college students in the new century are the generation that grows up with the Internet. They have natural sensitivity and affinity to big data related to the Internet. At the same time, the double pressure of employment and further study makes them have the consciousness to improve data literacy.

In the process of cultivating data literacy in colleges and universities, college students are both the training targets and the main force of data work and data technology research development

in the future society. Their data literacy directly affects the overall data literacy level of our society, and even directly determines the rise and fall of our data work in the future. Especially since the proposal of the 13th Five-Year Plan, big data strategy has increasingly become a basic strategy at the national level, and the strength of the future national competitiveness of big data is also reflected in the current training efforts and training efficiency of students' data literacy.

## 5. Conclusion

In terms of data literacy cultivation for college students, we should recognize and analyze the different subjects that constitute the cultivation model from the composition of data literacy, and build a multi-subject collaborative cultivation model from the practical point of view. University libraries, specialized teaching departments, off-campus information institutions and students form the four main bodies of data literacy cultivation in China's universities from four aspects: data awareness, data ethics, data skills and data practice. The four subjects have different needs for data resources, and also have the urgent desire and objective conditions for cooperation. Therefore, the cultivation of data literacy in colleges and universities should implement a four-pronged and multi-party cooperation program. In terms of cultivation, the combination of online and offline can be adopted to give full play to the role of the channel of the Internet. Focusing on the combination of theory and social practice, students' learning and the actual needs of government departments and social information institutions are given equal importance, so as to improve the efficiency and practice level of data literacy cultivation.

## Acknowledgments

This paper is the stage result of the quality project of Anhui Province Education Department (2020szsfkc0038); Research on users' perceptions and attitudes towards news algorithm recommendations" (SK2019A0519) of the 2019 Anhui University Humanities and Social Sciences Research Project.

## References

- [1] Z.A Zhang, Y.H Cao, Big data and news communication research hot spots and reflections. China Publishing, Vol.10(2017),3-11.
- [2] R.H Huang, B.Y Li, Data literacy education: the expansion of information literacy education in the era of big data. Knowledge of library intelligence. Vol.01(2016),21-29.
- [3] Hao, Yuan-Ling, Shen, Ting-Ting. The construction and strategic thinking of data literacy and its cultivation mechanism. Intelligence Theory and Practice.vol.01(2016), 58-63..
- [4] Li Qing, R.Y Ren. A review of the current situation of data literacy education research and practice of teachers abroad, Research on Electrochemical Education.vol.05(2016),120-128.
- [5] X.H Zhang. Comparative analysis of scientific data literacy education in university libraries at home and abroad. Librarianship Research.vol.11(2018),9-13.
- [6] L.J Deng, W.J Yang. Research on data literacy education in university libraries under big data environment, Library construction. No.01(2016),76-80.
- [7] Baidu Library. 2018 class training program of data science and big data technology:<https://wenku.baidu.com/view/b7333b0bf56527d3240c844769eae009591ba20f.html>.

- [8] China Social Science Network. Methods and strategies for data literacy education in university libraries:[http://ex.cssn.cn/glx\\_tsqbx/201709/t20170930\\_3658867.shtml](http://ex.cssn.cn/glx_tsqbx/201709/t20170930_3658867.shtml)
- [9] QQ.com.35 universities have successfully applied for big data majors: the shortage of related talents reaches 1.5 million:<http://edu.qq.com/a/20170331/021699.htm>
- [10]Xinhua.com. Xi Jinping: Implementing national big data strategy to accelerate the construction of digital China:[http://www.xinhuanet.com/2017-12/09/c\\_1122084706.htm](http://www.xinhuanet.com/2017-12/09/c_1122084706.htm)