ISSN 2616-7433 Vol. 6, Issue 6: 95-99, DOI: 10.25236/FSST.2024.060615

Three Philosophical Issues of Big Data

Haoran Meng[a], Fengya Gao[b,*]

School of Public Administration, Guangxi Minzu University, Nanning, Guangxi, 530006, China
[a]306316732@qq.com, [b]gaofengya439@163.com
[*]Corresponding author

Abstract: *Big data faces three philosophical issues. The first is digital ontology, which changes from static to dynamic, breaking the dichotomy between subject and object. The second is data cognition, which changes from the clear cognitive style of formula to the fuzzy cognitive style of construction and then from the mechanical superposition to the dialectical unity. The third is data reasoning, which shifts from the traditional science's emphasis on deduction to the big data era's emphasis on induction, and its function has changed from the balance of traditional scientific interpretation and prediction to the emphasis on prediction. These three philosophical issues reflect the phenomenological scientific philosophy tendency of big data. The entire information consists of subject construction. Nonetheless, data plays a role in each reproduction round, leading to a gradual increase in the construction factor of the subject. The starting understanding gradually shifts from precise to general, and the rough understanding as a whole is depicted in the general knowledge, leading to a comprehension of the object. In the age of big data, data is sourced from search engines, and design decisions are made using algorithms. The search engine's limits are predetermined, and users predefined the kernel based on their preferences and requirements.*

Keywords: *Big Data; Digital ontology; Data cognition; Data reasoning*

1. Introduction

The words digital, data, big data, cloud computing, Internet, and artificial intelligence are hot spots in academia today. Due to the development of the Internet, initially static and discrete data suddenly converged and transformed into big data, leading to unprecedented subversive changes in politics, economics, and culture. A harmonious universe is made up of numbers. The harmony of nature is the harmony of numbers, and the order of nature is the order of numbers. The categories proposed by Aristotle include objective existence in the category table and are divided into ten categories. Furthermore, when researchers analyzed the category of "quantity" in their first critique, they stated that the pattern of quantity is number. Social existence determines social consciousness, and Big Data's rise will inevitably entail related philosophical issues. This paper aims to discuss the five philosophical issues related to big data, including digital ontology, data cognition, data application, data reasoning, and data evaluation [7-8].

2. Digital Ontology: From Static to Dynamic

Data is an indispensable tool for human beings to explore the objective world. To be specific, data is an attribute of objective things. It not only records nature or human society but also presents various changes. Since the invention of numbers, humans have used numbers to record various data about nature and human society, from the Stone Age to the Big Machine Age. From the knotted ropes to record events, to Tycho's astronomical observation records, or even Newton's three laws and Einstein's theory of relativity, humans have recorded and represented things with numbers, data, and mathematical formulas.

Text and data are the most basic carriers of a civilization. Text is actually qualitative data. Through the text and data, such as the area of the region, the number of inhabitants, the speed of consumption, the output of production, the material support on which people's living and death depend, and the dependence of various resources to ensure the stability of regional rule are also a series of data [9].

Both nature and human society are inseparable from data. Philosophical significance lies in two points: One is quality; that is, numbers are the starting point of human reason. People must use numbers to understand natural objects and social objects. The second is quantity. The numbers and data characterize

ISSN 2616-7433 Vol. 6, Issue 6: 95-99, DOI: 10.25236/FSST.2024.060615

the measure of nature and human society. The primary way for the subject to understand the object is to grasp the object's data to obtain a sense of security and behavioral basis. Therefore, objective existences such as nature and human society cannot be separated from numbers and data. In this sense, data forms the basis of nature and social activity.

"People's existence shows their real-life process." [1] Marx's words reveal the essence of life. From the perspective of data, people's lives and their essence are constructed by data. "All natural scientists who act mathematically in their research make use of and must make use of metaphysical principles at all times (although not aware of it) [2]". Kant's words reveal the essence of natural science research. From the perspective of data, it reveals and explains the existence of natural objects and the mathematical structure of natural objects.

Data, from the beginning of the Paleolithic Age to the invention of the computer, existed in the above two ways. In addition, the philosophical hallmark is Claude Elwood Shannon's two papers: A Mathematical Theory of Communication and Communication in the Presence of Noise. The former article was published in 1948, and the latter article was published in 1949. These two articles laid the foundation for modern information theory.

If the data in the traditional era shows a static picture, the data in the era of big data is a dynamic data image like video. Since the publication of Claude Elwood Shannon's two fundamental papers, with the invention of computers, especially the Internet and mobile Internet, the world has gradually entered the era of Big Data.

Nowadays, people's lives are inseparable from big data. Whether watching the morning weather forecast when getting up, planning a long-distance route, booking tickets, high-speed rail, bus tickets, or leisure activities after a busy day, we are inseparable from big data. As for the work, we are inseparable from the computer and big data; even the air quality indicators should be concerned about. Data is available all the time and everywhere. Data determines the world, whether metaphysical or physical. The changing data flow determines the way and form of existence of this world at all times.

The most significant difference between modern and ancient data is that ancient data is discrete, independent, and static, while today's data exists in a constantly changing data flow. The movement of the objective world presents a series of figures representing change, which reveal the movement of social relationships and the change of social structure. Therefore, the modern world presents a picture of a dynamic digital ontology. Different things in the world present a variety of data, and similarly, different aspects of the same thing present a variety of data. It is a new ontology of digital process: it is related to phenomenology from the phenomenon level; from the essence perspective, it is interlinked with process philosophy, especially with Marxist practical philosophy. Therefore, it provides the possibility to bridge the gap between the subject and object of modern philosophy [10].

On the one hand, the opposition between subject and object in the traditional sense is dissipated in digital ontology, changing from static to dynamic. The objective movement of the material world is presented as a dynamic data flow in big data, breaking the single dimension and limited field of the subject-object dichotomy perspective. On the other hand, in the traditional sense, the opposition between subject and object is dissipated in the data source of the class that makes up the digital ontology. As the data source of the class, all data is interpreted as the transformation process from data element to data source. To sum up, it includes the initial state of the data element and covers the dynamic, changing, and development in the transformation process [11-13].

3. Data Cognition: From Certainty to Approximation

The number is a concept. The data is deeply rooted in people's hearts because of its certainty and clarity. Digital resources are infinite, and unique data can represent all things in nature and society. In this sense, the Pythagoreans declared: "Everything is countable."

Since the birth of human society, human beings have always sought certainty. Data spans time and space, representing certainty. During the operation of human society, population statistics are done with data, and land measurements are done by data, data determine national territory, labor results are statistics by data, and even statistics on losses caused by natural disasters are expressed in data. Data is used for no other reason than its clarity and precision. Examining human scientific research, such as the Descartes-Newton research method, with the help of scientific experiments, the objects of nature are split and decomposed into data. Then, the data is restored to the whole research object again. Finally, the cognition of the object is obtained. Therefore, research is inseparable from the clarity and accuracy of data.

ISSN 2616-7433 Vol. 6, Issue 6: 95-99, DOI: 10.25236/FSST.2024.060615

From the perspective of philosophy, the approach of traditional cognition is the theory of active reflection of subject to object. The object is objective, and the representation data must be objective to accurately represent the object. Data is used as an intermediary bridge because there is a natural gap between the subject and the object. Because each piece of data is unique, the object it represents must be clear and precise. The philosophical foundation of representation is based on substantive thinking and small data. Because data are based on various angles, they are discrete, separate, and static. However, these data are all reflections of the actual attributes of the entity and can restore a clear picture of the entity from different aspects [14].

There is a view that the definition of data's own attributes (such as quantity and type) is too vague, resulting in the ambiguity of big data, so big data is in a "dangerous situation of existence and nothingness". [3] For philosophy, this paper argues that the so-called big data subverts the cognitive model of traditional data concepts. The fundamental feature of big data is that the continuous circulation of data leads to data reproduction, which fundamentally subverts the subject-object dichotomy model of traditional cognition and produces a new cognitive model: the subject-object fusion cognitive model. As a result, the boundary of cognition is gradually blurred.

Let's start with data cognition of natural objects, which is full of continuous data flow. That is to say, the cognitive objects of nature are represented and constructed in the data flow. Subject and object are constantly becoming mixed and blurred [15]. Authenticity and fuzziness coexist, and the boundary is blurred. In other words, the cognitive object of social objects is characterized and constructed in the fuzzy data stream: there is virtuality in reality.

When talking about the data cognition of social objects, the objects are full of data flows that are difficult to distinguish between real and fake. More and less, true and false coexist, and the boundaries become increasingly blurred. That is to say, the cognitive object of the social object is constructed in the fuzzy data flow: there is virtual in reality, and there is reality in virtual. As a result, the cognition of big data transforms from clarity to approximation.

Philosophically, there are two ways to understand big data. One represents natural objects or social objects through data flow, and the other is to construct natural objects or social objects through data flow. However, it isn't easy to completely separate the two. They intertwine and penetrate, constantly in a state of inseparability [16]. The purpose of making a distinction is to facilitate our discussion in theory, but in practice, it isn't easy to distinguish.

All the data has the composition of subject construction. However, one by one, data is involved in the reproduction round by round, and the construction factor of the subject is gradually enlarged. First, the initial cognition is inevitably from clear to approximate, and its overall picture is described as a whole in the approximate cognition, so we get a cognition of the thing. Second, in the era of big data, data comes from search engines, and design is based on algorithms. The boundary of the search engine is set in advance, and the user sets the kernel in advance according to their preferences and needs. After a series of algorithms, the search engine pushes the information that users need. At the same time, users no longer pursue absolute clear goals but look for approximate fuzzy results under macro trends. This cognitive approach departs from the pure objectivity of traditional data and shows traces of subjective penetration. However, the way of information filtering is objective. The above is the paradox of big data, and the approximate results are obtained in an accurate way.

4. Data Reasoning: From Deduction to Induction

The essence of data is the essence of people. Whether natural or social objects, they exist as objective forms of people's cognition, and data representation confirms the results of cognition. From the data described by the starting point to the object, to the data revealing the object's internal structure, and finally to the data representing the cognitive results, what kind of reasoning path has the data gone through?

There are two traditions in modern science: classical science and empirical science. The former includes mechanics, optics, astronomy, and mathematics, with mathematics as a rational tool. The latter focuses on qualitative and quantitative measurements, using instruments, emphasizing experimental results, and using data as a rational tool. Classical science focuses on deduction, and empirical science focuses on induction. But both are inseparable from the scientific hypothesis, deduction, and reasoning. In classical science, deduction and reasoning are played to the extreme, and Newton's law is an outstanding representative. Einstein said: "Newton succeeded in explaining the motions of the planets, satellites, and comets down to the smallest details, as well as the tides and the movement of the earth - a

ISSN 2616-7433 Vol. 6, Issue 6: 95-99, DOI: 10.25236/FSST.2024.060615

deductive achievement of unique magnificence." However, in empirical science, people establish a scientific model based on the experimental results, put forward a scientific hypothesis, verify it through experiments, and finally put forward a scientific theory. Deduction and reasoning are also very important for the "empirical science" of hypothesis-experiment-confirmation-theory.

From a philosophical perspective, traditional science assumes the dual responsibility of explaining and predicting the world. Therefore, traditional science's reasoning methods include both inductive and deductive reasoning [17]. This is the reasoning path of the scientific methodology of rationalism in modern times. "Scientific methodology is structured based on testing hypotheses. Most of these scientific models are implemented in the mental system of scientists. The models are tested, and experiments confirm or falsify the theoretical models of how the world works. That is how science has worked for hundreds of years." Since modern times, due to the theory-driven research paradigm, deductive reasoning has made significant scientific achievements. Sound, light, electricity, heat, magnetism, biology, and chemistry are all subject to deduction and reasoning. For example, Dalton's chemical system, Rutherford's atomic system, Coulomb's electrostatic formula, Ampère's electric force formula, and Mendeleev's periodic table of elements all applied this reasoning method, not to mention Newton and Einstein.

However, the emergence of the big data era has disrupted this theoretical model of scientific research. The numbers can speak for themselves as long as there is enough data. "Everyone besides God has to talk in data." [4] Data in the above fields are all one-dimensional objects of scientific research, technological inventions, engineering and construction, economic activities, and social life. However, big data forms a 360-degree global record of the object [5]. Big Data means there are no blind spots, and the subject matter is comprehensively represented through big data, providing an excellent basis for inductive reasoning. As long as the amount of data is nearly complete, the amount of object features is also roughly complete [6]. The various aspects of the data confirm and support each other. Big data provides a solid foundation for inductive reasoning to complete the prediction of scientific research. To improve the accuracy of big data analysis, researchers should pay attention to algorithm design and implementation. Big data analysis is closely related to artificial intelligence, and there are many technical similarities, such as the need for model design. Modeling is also essential for big data analysis. In the cloud computing era, the practicality of machine learning has dramatically improved due to improved computing power and increased data volume, especially in deep learning. Analysts must have a certain level of industry knowledge to increase the application value of big data.

5. Conclusion

Deduction and induction are both essential parts of scientific practice and have greatly promoted the progress of science and technology. It is a new trend in scientific research that successful "prediction" is based on inductive reasoning. It focuses on inductive reasoning, abandoning scientific explanation's function and retaining scientific prediction's function. The person who mastered the data will master everything. A new empiricism came into being in the era of big data. Deductive reasoning based on traditional scientific theoretical models is not always absolutely reliable. However, in the era of big data, people do not directly base scientific research based on inductive reasoning with massive data. In philosophy, the reasoning path of big data is consistent with the phenomenological philosophy of science that has become popular in the West in recent years. Various data replace phenomena associated with objects, and phenomenology is digitized, especially surrounded by big data.

References

[1] Marx K, Engels F. Marx & Engels Collected Works [C]. Beijing. People's Publishing House. 2012.
[2] Kant. The Complete Works of Kant [C] (Vol 4). Beijing. China Renmin University Press. 2005.
[3] Simsek Z, Vaara E, Paruchuri S, et al. New ways of seeing big data[J]. Academy of Management Journal, 2019, 62(4): 971-978.
[4] Einstein A. The general theory of relativity[M]//The Meaning of Relativity. Dordrecht: Springer Netherlands, 1922: 54-75.
[5] Anderson C. The end of theory: The data deluge makes the scientific method obsolete[J]. Wired magazine, 2008, 16(7): 16-07.
[6] Tu Z P. Big data: The coming data revolution[M]. Guangxi Normal University Press, 2013.
[7] Dong C Y, Xue Y H. The epistemological value of personalized knowledge in the big data era [J]. Philosophical Trends. 2018 (1): 95-101.

ISSN 2616-7433 Vol. 6, Issue 6: 95-99, DOI: 10.25236/FSST.2024.060615

[8] Huang X R. Big data, digitization, and scientific boundary [J]. Dialectics of Nature Communications. 2018 (5): 118-123.
[9] Yu, Xie Z Q. Metaphysical foundation in the era of big data [J]. Zhongzhou Academic Journal. 2015 (5): 106-110.
[10] Wu C W, Ren S Q, Deng. Analysis of the view of big data from the Marxist philosophy of technology [J]. Qiushi. 2017 (7): 4-14.
[11] Wu J, Zhai T F. Big data and epistemology [J]. Philosophical Research. 2015 (11): 110-113.
[12] Zhang G H. On the nature of data and its relationship with information [J]. Philosophical Analysis. 2018 (2): 119-132.
[13] Chen Y, Li J L. Review of Big Data Philosophy Research [J]. Guangxi Social Sciences. 2017 (3): 50-55.
[14] Fang H F. Big data: history, paradigm and epistemological ethics [J]. Zhejiang Social Sciences. 2015 (9): 113-120.
[15] Su Y, Wei Y D. The mechanism and significance of big data knowledge representation [J]. Research in the Philosophy of Science and Technology. 2017 (2): 63-68.
[16] Ma Y. Big data and human development [J]. Philosophical Analysis. 2018 (1): 105-115.
[17] Chen M, Mao S, Zhang Y, et al. Big data applications[J]. Big Data: Related Technologies, Challenges and Future Prospects, 2014: 59-79.