

大数据的存在论意蕴

王天恩

摘要 作为信息数字编码发展的产物,大数据扑面而来并整体呈现出一个不同于物能世界的信息世界。作为不同于物能存在的存在,大数据具有重要的存在论意蕴。由于越来越以数据网络的形式存在,且通过使物能存在网络化,大数据意味着存在的升级。这种存在升级意味着存在由物能世界到数据世界的扩展,信息和物能存在的网络一体化,存在过程及其把握的更高维度的提升,人和世界存在与发展信息层次的一体化展开。基于信息存在的大数据形态,不仅可以得到“ontology”涵义从本体论到存在论嬗变的大数据观照,看到从本体论到存在论、从实体存在论到关系存在论、从最基本的物的存在到最高层次的人的存在、甚至从存在到非存在嬗变的逻辑,而且可以看到存在的未来已来维度,从而进一步看到大数据基础上存在论与人类生存和发展的更深层次关联,直至存在论的人类学意蕴。

关键词 大数据;ontology;信息编码;存在论意蕴

中图分类号 B086;TP39 **文献标识码** A **文章编号** 1672-7320(2022)02-0029-09

基金项目 国家社会科学基金重点项目(17AZX003)

作为一个具有规模整全性、实时流动性、结构开放性和价值生产性的量化世界^[1](第15版),大数据越来越清晰地晓示着自己非同寻常的存在论意蕴。通过大数据呈现的量化的信息世界,可以看到存在的整体扩展和升级。作为信息数字编码发展的产物,大数据的发展在物能世界的基础上,逐渐呈现出一个越来越具有整体性的信息世界。正是在物能存在基础上信息存在的发展,意味着存在的升级。通过信息存在的大数据形态,从大数据层次观照“ontology”的涵义嬗变以及与人类生存和发展更深层次的关联,可以看到大数据的重要存在论意蕴。

一、大数据构成的存在升级

大数据构成的存在升级,一方面表现为存在的大数据内涵扩展,另一方面表现为存在的大数据维度提升。在人类文明发展之初,人们的活动主要基于大自然,在与大自然的相互作用中主要通过理解世界来理解自己;随着数字技术的发展,人类可以在大自然的基础上,在更高层次进一步基于大数据活动,在与大数据的相互作用中通过理解自己来更好地理解世界,从而在更高层次构成了人类理解世界和理解自己的双向循环。基于自己活动结果的大数据来理解自己,正是作为人类文明发展伟大转折的信息文明的实质。正是信息文明的发展,徐徐展开了大数据丰富深刻的存在论意蕴。

信息科技特别是数字技术的发展,逐步形成了一个只具有存在论预设的大数据世界。虽然大数据和大自然性质截然不同,但就人类而言,大数据只具有存在论层次的在先预设,意味着大数据在基本方面与大自然具有存在论对等性。这既是大数据最重要的哲学意蕴所在,也是大数据区别于小数据的根本之处。大数据和大自然在基本方面的存在论对等性,使我们可以在与大自然存在论意蕴的同等意义上讨论大数据的存在论意蕴。大数据不仅集聚了人们在网络上留下的数据足迹,而且将相对碎片化的

数据足迹整合为一个关联丰富的信息世界。数据足迹不仅使我们能建立起与空间立体相关的关系结构,而且能够在时间维度构成历史、现在和未来的动态相关关系。当由此把握得到的大数据与世界同步流动甚至可以快进和慢动作考察时,我们就在物能世界的基础上,拥有了一个与自然世界具有基本层面的存在论对等性,甚至比自然世界具有更丰富可能性的大数据世界。

在存在论层次,大数据存在一方面在基本层面具有与大自然的对等性,另一方面又具有大自然所没有的更重要的性质和特点。这些不同的性质和特点主要表现在三个方面:一是作为信息数字编码发展的产物,大数据是一个量化的世界,集中体现的就是大数据“量化世界”的口号;二是作为人类活动的结果,大数据具有人类学特性,大数据(表明信息存在具有客观性和非确定性相结合的两方面)不仅进一步明确了量化把握和质性把握的性质及其相互关系,而且表明了信息领域客观确定性和主观非确定性关系的更深层次,其极端表现就是大数据的可篡改性;三是在大数据基础上未来纳入当下形成的更大程度可预测性,典型标志就是“未来已来”。大数据的存在性质及其新的存在特点,无疑具有重大存在论意蕴。与大数据的规模整全性一样,大数据的存在论意蕴,既是指具体的大数据,更是在大数据的融会发展意义上说的。大数据的发展必定会经历一个数据割据的发展阶段,但统一发展也必定是大数据发展的总体方向。大数据越是统一发展,其存在论意蕴越典型。正是由此才可以看得更清楚,大数据构成了存在的升级以及新的存在论性质。

大数据构成的存在升级,具有十分丰富深刻的内容。

首先,大数据构成的存在升级表现为物能世界基础上数据世界的发展,意味着存在的扩展。作为信息数字编码发展的产物,大数据以越来越具有整全性的方式,使存在从物能存在扩展到信息存在。正是作为整全的数据存在,大数据才具有信息意义上的存在论意蕴。在最基本的意义上,“信息是信息,不是物质或能量”^[2](P132),而是一种关系,一种感受性关系,一种基于物能的感受性关系,其成熟形态即信宿和信源间的感受性关系^[3]。而作为机器信宿与信源相互作用的结果,信息的数字编码是自然生成的,类如自发生成的大自然。如果大自然是原生存在,那么,大数据就具有次生存在的性质。由于还没有进入人类认识,大数据的这种次生存在不是大自然存在的延伸,而是既不同于物能存在,同时又具有世界存在性质的信息存在。这个以信息方式存在的世界,正是在量化世界中生成的新的世界层次。

随着数字化的发展,关于大数据“量化世界”的口号,一方面形成越来越普遍的共识,另一方面又存在理解上的分歧。在法国哲学家拉图尔看来,量化世界具有明显局限性。以社会学为例,由于在社会领域中划分出了什么是、什么不是可量化的知识,社会学从未成为一门量化的科学。拉图尔以法国学者特有的风格写道,“数字、数字、数字,社会学一直被成为一门量化科学的目标所困扰。”^[4](P145-162)根据拉图尔的观点,微软研究院首席研究员、《数据与社会》的创立者达纳·博伊德(Danah Boyd)和克劳福德(Kate Crawford)认为,“大数据为人文学科提供了一条新的途径,以确立量化科学和客观方法的地位。它使更多的社会空间可以量化。在现实中,基于大数据的工作仍然是主观的,它所量化的东西并不一定能证明客观真实——特别是在考虑来自社交媒体网站的消息时”。在他们看来,“声称客观性和精确性是误导”^[5](P4)。在这里,明显存在一个关于量化世界的更深层次的理解问题。由于数据化意味着量化,大数据的口号就是量化世界;由于算法不仅有数值运算算法,还有非数值运算算法,而非数值运算算法几乎可以无所不包或没有确定的界线,这个意义上的量化触角的确可以伸向几乎任何领域。在主流观点看来,数据化“指的是获取有关天下万物的信息——包括那些我们过去根本不认为是信息的东西”^[6](P15),因此问题的关键在信息。只有在信息层次才能看到,作为量化的世界,大数据是一个信息数字编码的世界。作为由既不是物质也不是能量的信息构成的世界,作为量化的世界,大数据不仅构成了一个与物能世界具有基本层面存在论对等性的存在,而且其构成的一个具有不同存在论意蕴的全新世界。由于信息是物能基础上涌现的感受性关系,作为信息数字编码发展的产物,大数据不仅意味着“存在”范围的扩展,而且意味着“存在”基本层次的提升。

其次,大数据构成的存在升级意味着信息和物能存在的网络一体化。这种存在升级,不仅意味着在由物能世界扩展到作为信息数字编码的数据世界,更表现在存在方式前所未有的提升——网络化。正像信息基于物能、大数据基于大自然,相对于大自然,大数据具有一个重要优势:大数据的网络化几乎就是其天然本性。大数据的特性与数据规模密切相关,而网络则是大数据规模扩展和层次提升必不可少的基础。随着网络化的发展,大数据越来越以数据网络的形式存在。这意味着,大数据的网络化一方面以信息的网络化为基础,另一方面又不仅使信息网络化,而且为物能的网络化提供了更高层次的基础。

大数据构成的存在升级,也不仅意味着信息世界的网络化,而且表现为信息和物能存在的网络一体化。正是由于大数据最基本的存在论意蕴是从物能存在到信息存在的扩展和提升,在大数据基础上,不仅信息存在,而且物能存在都在走向网络化。大数据典型地晓示了以人的需要为出发点,以人的需要的满足为最终目的的哲学本性。作为创构的基础,大数据不仅为创构认识论^[7],而且为具体而非抽象的存在论研究提供了新的前提性条件。在这个意义上,大数据不仅启示了更深层次的信息知识论,并为算法存在论^[8](P35-59)研究提供了存在基础。也正是在这个意义上,大数据提供了超越原来方法论划分的局限,在存在论和知识论发展一体化的基础上,不仅提供了实现方法论新突破的条件,而且促进了更具存在论意蕴的信息和物能一体化,这在信息和物能网络化融合的物联网整合中表现得最为典型。

在大数据和云计算等信息科技发展的基础上,从互联网到移动网络再到物联网的发展,使信息世界和物能世界逐渐实现无缝对接,世界在数字化基础上重新整合为一个信息基于物能并整合为一体的全新世界。其核心机制就是万物互联的物联网。

物联网由智能终端相连构成,智能芯片作为网络节点将万物互联。在物联网构成中,射频识别(RFID, Radio Frequency Identification)系统有读写器(Reader)和电子标签(Tag),其作用是在读写器和电子标签之间建立起感受性关系。电子标签既是感受器也是信号发射器,它从读写器的电磁波获得能量,把所标注的物品的信息编码发射出去,使读写器与之建立起关于该物品的感受性关系。由此可以建立起物联网,通过每一件物品上的电子标签将物品与互联网联系在一起,也就是通过根据需要建立物品和物品间的感受性关系,实现物能之间联系的信息化。而移动的网络终端则使泛在的物联网日益摆脱特定空间格局限制,构成了一个超越固定空间、信息物能一体化的大数据世界。物联网使作为信息文明基础的大数据与物能基础渐趋无缝连接,正构成与物能文明有机结合在一起的信息文明基础。云计算则以典型信息及其处理共享的方式,极大地强化了单纯大数据存在的割据状态,为信息文明基础的彻底整体化奠定了基础。由于物联网意味着信息和物能的网络一体化,在主要是数据之间相关关系的基础上,不仅生长出越来越丰富的物能关系,而且必定使大数据相关关系和因果关系之间发生重要的关系性变化,这种变化具有重要存在论意蕴。

再次,大数据构成的存在升级,意味着存在过程及其把握的更高维度提升。大数据构成的存在升级,甚至不仅意味着物能和信息存在一体化意义上的网络化,而且表现为存在过程及其把握的更高维度提升。由此,存在不仅出现量的扩张,而且发生了重要增维。信息存在不仅是纯粹物能意义上的存在扩展,而且由于物能和信息日益深度融合,具有实时流动性的大数据不断向未来维度延伸,使大数据存在具有越来越强的预测功能。预测是大数据的核心,正是大数据的预测功能,使存在空前向未来维度扩展。由于信息不同于物能的性质,大数据整体呈现的信息存在使整个存在状态的涵盖范围越来越朝未来向度伸展。随着人工智能的发展,可以看到未来向度延伸日益充分的展开。这既与信息的网络化优势、又与信息的本性密切相关。

最后,大数据构成的存在升级,意味着人和世界的存在发展信息层次一体化的展开。大数据构成的存在升级最重要的还在于,这不仅意味着物能和信息一体化存在过程及其把握的更高维度的提升,更表现在大数据提供了一个新的平台,在此基础上,越来越主要作为信息方式存在的人建立起了与自然世界更深层次的内在关联,从而展开了人和世界存在发展的信息层次一体化。网络不仅使大数据有了广泛

来源和数据多样性,更重要的是使大规模数据集构成了丰富的相关关系。这是人作为社会关系的总和,人类社会构成信息层次关系性升级存在的重要内容。随着信息科技的发展,网络越来越成为人类生存的直接基础,数据的由此产生和由此产生的数据,就与抽样采集的数据具有原则性不同。如果说,作为抽样采集的数据,样本数据是资料性数据,那么,作为人类生存的数据足迹整体,大数据则是具有存在论意义的数据。正是由此,才构成了具有规模整全性、结构开放性、实时流动性从而价值生产性的大数据,构成了人类存在和发展越来越必不可少的基础。正因为如此,大数据才不仅具有世界观意义,而且通过信息的数字编码,在物信息化和信息物化的直接关联基础上,构成了人和世界存在发展过程的双向循环一体化机制。在这一机制中,大数据存在的网络化层面和未来维度具有前所未有的存在论意蕴。

从主要由过去和当下构成的存在到未来已来的存在,从相对离散的存在到存在的网络化,从互联网到物联网,从物能存在到信息存在,人和世界的存在性关联从物能层次到信息层次,大数据世界提供了一个重要基础,构成了信息存在的大数据形态。

二、信息存在的大数据形态

由于大数据构成了一个越来越类似物能世界的信息世界,对人类来说,大数据本身的存在形态在某些方面就表现得越来越类似大自然。关于大自然,理解可以很复杂。广义上的理解,人类也可以包含在大自然之内,要进行更有意义的区分,则从无机物到有机物,从植物到动物直到人类,不可能有明确的界线。至少相对于人类而言,信息和物能的原则区别为我们更合理地使用“大自然”概念提供了更深层次的依据。作为感受性关系,信息的产生及其物能编码特别是观念编码的存在,不仅构成了一个与大自然区别越来越大的世界,而且这种区别比自然世界和人工世界的区分更具合理性。在很大程度上,这与信息编码的数字化发展密切相关。

信息是以其编码进入操作的。信息编码有两种基本类型:信息的物能编码和信息的观念编码。前者典型的如作为信息物理编码的电信号;后者典型的如作为可感具体事物抽象概括结果的概念。在典型的物能编码和典型的观念编码之间,存在一个连续系列,其间信息的数字编码具有特殊地位,在两种基本信息编码之间起着至关重要的直接转换作用。在数字编码基础上,信息的物能编码和观念编码可以近乎完美地融为一体,这不仅在物能和信息关系中意味着二者在机制上的一体化,而且在人和世界关系中意味着更深层次的内在关联。由此可见,深化信息数字编码理解的意义以及关于大数据及其存在论意蕴研究的重要性。在信息意义上,大数据正是这样一个可以和大自然具有基本方面存在论对等性的世界。在信息世界和人的关系中,出现了全新的内容,与人和自然世界的关系完全不同。这一具有重要意义的发展,与信息存在的大数据凸显密切相关。

作为信息数字编码发展的产物,大数据空前开显了信息。以信息、大数据和人工智能为主要标志,信息科技的发展构成了具有生态维度的“信息文明之树”。大数据就像信息文明之树的树干,信息就是树的根,人工智能则是信息文明之树的花果。大数据和人工智能是信息的展开。大数据以信息数字编码的方式,将生长在物能土壤中的信息伸展出物能大地。以大数据为典型代表,信息科技的发展使人们对信息的资源性质认识越来越自然化。人们甚至认为,“信息是我们这个世界运行所仰赖的血液、食物和生命力”。而比特“存在于一个个二进制数字、一个个触发器、一个个‘是’或‘否’的判断里”^[9](P5,7)。如今,通过展开信息,大数据以信息数字编码的方式逐渐将信息和物能直接关联在一起。

大数据的发展前所未有地展开了信息。信息完全不同于物能的本性,人类借助信息技术处理数据,越来越可深达存在论层次。不仅如此,回过头来由信息看物能,使人类的物能处理能力随着信息处理能力的发展而提升,从而物能观念也相应发生改变。随着信息和大数据技术的发展,由于人类可以通过信息驾驭物能,依靠人工智能处理物能也会有一个根本不同的前景。正是与物能存在的这种内在关联,使信息成为对人类来说越来越重要的存在。大数据作为信息的展开,正形成信息存在的大数据形态。

作为技术,大数据的意义可以想见;但作为信息性存在,大数据的发展前景至少到目前我们还难以想象。关于大数据在信息存在中的地位,有一种表述是“大数据走向信息世界的中心”^[10](Pxxv)。这意味着,大数据最重要的意义不在大数据方法,而在其存在论意蕴。“在比我们想象中更近的将来,我们世界的许多方面都将被计算机系统增强甚至取代……正如互联网通过将通信加入计算机根本改变了世界,大数据也将这样,通过给予世界一个从未有过的量化维度,大数据将改变生活的基本方面。”^[6](P12)大数据提供或更确切地说所开辟的“可量化的维度”,其实正在成长为一个世界,或至少是世界的一个离人类最近的层面。对人类来说,它就是一个虽然是量化但同样具有整体性的世界,即一个量化的整体世界。它一方面具有与自然世界基本方面的存在论对等性,另一方面又具有自然世界没有的性质。正是因其特点,大数据量化整体世界深层次涉及大数据与物能和信息的关系。

关于大数据和信息的关系问题,与对信息的理解密切相关。如果明确了信息是感受性关系,那么一切可感受的存在就都能与信息建立起联系,甚至包括“非存在”。在这里,非存在不是与抽象的存在相对的抽象观念,而是与感性具体事物的存在相对的非存在。比如窗台上一盆花的不存在没有物能意义,但在信息意义上,它可以是与之建立起重要信息的信源。在量子通信中,“反事实通信”正是以具体信号的非存在为信源进行的。在物能意义上,有意义的非存在只能是对应的存在观念;而在信息意义上,非存在则可以是与存在一样具有实践甚至操作意义的存在。观念不能作为可以与之建立起感受性关系的信源,因为观念不具有感受质(qualia),可以作为信源的非存在是感性具体存在的反面。正是作为感性具体事物存在的反面,非存在可以是与信宿建立起感受性关系的对象。所有与感性具体存在对应的反面都可以是潜在信源。这不仅意味着存在范畴新的拓展,而且意味着存在观念的重要更新。就感性具体而言,不仅具体事物的存在,而且其非存在,都与感性实践具有根本关联。由此可以在更深层次看到大数据作为信息存在形态的存在论意蕴,看到这个意义上大数据存在论意蕴的非同寻常。

人类文明伊始,数据就伴随着人们,大数据则空前凸显了数据的存在。大数据存在的特殊意义,无论在哲学还是科学中,都引起了前所未有的关注。在李德毅院士看来,“大数据本身既不是科学,也不是技术,它反映的是网络时代的一种客观存在”^①,数据存在的大数据方式,空前凸显了不同于物能存在的信息存在基本事实。“这些趋势的更深层次原因是我们有更多的数据。而我们拥有更多数据的原因则是我们正在以数据形式呈现现实的更多方面”^[6](P72)。在这里,“以数据形式呈现现实的更多方面”,既意味着能看到越来越多事物的数据化存在方式,也意味着数据的存在越来越普遍。对大数据来说,这方面具有重要意义的一个更到位的表述,应当是数据——事实上在根本上说是信息——成了越来越重要的基本存在。正是在这个意义上,大数据具有异乎寻常的存在论意蕴,它不仅整体性地凸显了信息存在这一越来越重要的全新基本事实,而且更重要的是将物能存在提升到一个新的层次。所有这一切,都与大数据作为信息数字编码的数据集合密切相关。

作为信息数字编码发展的产物,大数据中的数据是信息的展开。看不到这一点,关于大数据甚至数据的理解就会有根本局限。《大数据时代》一书影响很大的原因之一,就在于迈尔-舍恩伯格等建立在技术基础上的哲学眼光,这种哲学眼光的重要表现之一,是看到作为大数据背后的信息:“大数据发展的核心动力来源于人类测量、记录和分析世界的渴望。信息技术变革随处可见,但是如今的信息技术变革的重点在‘T’(技术)上,而不是在‘I’(信息)上。现在,我们是时候把聚光灯打向‘I’,开始关注信息本身了。”^[12](P97)迈尔-舍恩伯格等不仅关注到大数据的信息关联,而且开始了大数据的信息探索。正是深入到信息层次,不仅使信息存在得以开显,而且大大改变了物能存在的状态以及我们关于物能的理解。信息存在使我们对物能存在的认识提升到一个新的层次,从而为存在论本身理解的深化提供了新的时代条件。这些新条件构成了一个很大的探索空间,其中蕴含着存在论范式的重要转换。正是信息存在

① 李德毅院士在“新一代信息技术产业发展高峰论坛”上的演讲^[11]。

的大数据形态,为深入理解“ontology”概念涵义的演变提供了更高层次的整体观照。

三、“Ontology”涵义嬗变的大数据观照

作为贯穿整个哲学发展史的基本概念,“ontology”涵义的嬗变本身成了一个重要研究课题。从本体论到生存论,从实体存在论到关系存在论,从最基本的物的存在到最高层次的人的存在,甚至从存在到非存在,无不与“ontology”密切相关。由于涉及不同于物能存在的信息存在,大数据为理解“ontology”涵义的嬗变提供了信息层次的整体观照。

由于数据尤其是数字数据越来越重要的地位和意义,深化数据尤其是大数据的存在论理解至为重要。“在拉丁文中,‘data’(数据)一词在‘事实’意义上指‘给定’(given)。”^[61](P78)这和哲学中将经验看作给定的事实具有内在关联,其存在论意义不言而喻。在这方面,科学和哲学界都已有不同程度的关注。徐之沛分析了数据、信息和知识的区别和联系,提出了数据、信息和知识三层次结构:数据是基础,在数据中提取信息,将信息加工成知识^[13](P88)。作为信息编码最具信息性质的形态,数字数据的确十分特殊,与给定的经验一样具有存在论意义。数据肯定是给定的事实,其以数字方式内含的是不同于物能事实的信息事实。如果说,物能事实意味着实在,那信息事实则既不是物能意义上的实在,亦非数学和逻辑意义上的实在,数据在任何意义上都不具有经典实在论涵义,而只具有事实存在的存在论涵义——这正是具有本体意味的实在和不具有本体意味的具体存在的区别,也就是实在论与存在论的区别。实在论一定具有本体承诺。在传统形而上学中,实在论承诺的是通过抽象得到的终极形式,无论是经验的还是逻辑的。由于是通过抽象得到的概括结果,属于未经检验的基本预设。存在论与实在论的根本区别在于,存在既可以是实在论意义上的存在(往往是抽象的存在),也可以是具体事物的存在,作为过程就是具体的事件。因此,实在论是从本体论的退离(back away),存在论又是从实在论的退离。本体论、实在论和存在论,三者不是取代而是范式发展递进的关系。

本体论和实在论都是纯粹的基本预设。由于本体论涉及抽象的终极实体,在一些研究领域,其合理性已经退离到实在;实在性之所以不可被否认,也因为可以指具体事物的实在性。存在论则退离到可以完全没有抽象的终极实体承诺,只要有具体事物的存在,就有自己的合理性;这是人类理性不能否定的,任何主观否定都将失去合法性。也就是说,存在论可以包含抽象的存在和具体的存在者。存在不能没有存在者,存在者却可以没有存在,正像“水果”不能没有苹果、香蕉和桃子(其实这里连“苹果”和“香蕉”等都不能用,只能用指物定义)这类具体事物,而这类具体事物却完全可以没有“水果”。只是人类需要这些抽象概念,否则就不可能有真正意义上的人类文明,甚至人类不可能走出蒙昧状态,不可能有人的开化。由此可见存在论在信息层次特有的合理性;由此也可见,只有在深入信息层次的大数据中才能进一步明确这些概念的各自意义。存在论具有本体论甚至实在论不具有的概念优势。实在论相对于本体论的优势在于实在可以退离到数学实在等,而存在论相对于本体论和实在论的优势则在于它甚至可以退离到非存在。这种合理性在信息层次可以看得更清楚。作为信息编码的数字数据就最典型。

与模拟数据不同,关于数字数据,我们既不能理解出原初意义上的“本体”涵义,也不能理解为数学意义上的实在,而只能在具体存在意义上理解。因为数字数据作为既非物能亦非数学或逻辑的信息编码,只能理解为具体的存在。在这个意义上,大数据与波普所说的“第三世界”显然不仅有关联,而且“大数据具有与世界1和世界2平起平坐的地位”^[14]。但在存在论意义上,这种平起平坐是就物能对信息而言的。在这个意义上,大数据则既有与“世界3”平起平坐的基本存在论对等性一面,又有与信息和本性上不同的一面。正是由于这种不同,特别是大数据未来已来的维度展开,可以通过知识维度和未来向度,在信息层次走向大数据存在论意蕴的人类发展关联。

存在主义的本体论研究将抽象的存在拉回到具体的人的存在,大数据则提供了一个在信息层次不仅深理解存在论,而且深理解具体的人的时代条件。大数据存在论意蕴的未来向度和知识维度,归

根结底将在信息层面构成人类生存和发展的存在论涉及。作为生物进化到最高阶段的产物,大自然主要在生物意义上孕育了人;作为基于物能发展的信息体,大数据的发展则给人提供了数据襁褓。“我们被包裹于数据襁褓。……我们与这些数据并不像我们所想象的那样可分离。……我们的信息构成了我们自身的重要部分。……就像任何襁褓,我们的数据既让人感到约束,又让人感到欣慰。信息开启了我们成为什么样的人的可能性。然而,它也排除了我们不能成为什么的可能性。……我们学会生活在我们的信息格式中,并最终依赖于它们。”^[15](Pvii)作为具有与大自然基本方面存在论对等意义的存在,大数据在人的存在和发展中的地位可想而知。的确,不同的人群会感受到大数据不同的机遇和挑战,但大数据基础上的创构以人的需要及其发展为出发点,以满足人的需要及其发展要求为最终目的的根本性质,决定了大数据存在之于整个人类存在和发展的重要意蕴。大数据的发展,一方面和几乎所有科技重大发展一样,必定因受益程度不同而扩大人与人之间的差别,另一方面又在归根结底的意义上为人类发展到更高层次提供前所未有的关键条件。作为既定的自然存在,大自然不内含存在论意义上的人类因素,而作为人类活动的产物,大数据存在却通过智能算法等内在包含人类的价值观等,这无疑意味着存在性质的颠覆性转换,必定使存在论在更深层次涉入人类的生存和发展,既构成人类生存和发展的历史性机遇,又构成新的严峻挑战。由于大数据新的存在特性,在人类的生存和发展中将出现一些具有挑战性的全新存在论问题,其中最典型的的就是大数据篡改。

作为大数据不同于大自然的性质和特点,大数据的可篡改性与大数据由人类活动形成、以信息的数字编码方式存在密切相关。大数据篡改不仅是最容易发生而且是最难以被发现的事实篡改,而且由于大数据具有存在论意蕴,它还是具有存在论意义的篡改。这是最基本也是影响最大、最严重的篡改。数据一变,大数据相关关系相应改变;大数据相关关系的改变,可以意味着相应规律的变化。更为重要的是,大数据篡改以极端的方式表明了大数据存在论意蕴一定程度的属人性。作为既定的物能存在,大自然没有存在论意义上的篡改问题,即不存在通过改变自然影响自然法则的可能,而作为人类活动的产物,大数据却存在具有存在论意义的篡改问题。这不仅意味着大数据存在论意蕴的人类学性质,而且对人类生存和发展构成了具有存在论意义的根本挑战。由于大数据产生于人类生存活动,大数据的发展和人的发展将构成具有存在论意义的关联。在这一存在论关联中,数据挖掘具有特殊意义。

通过数据挖掘,大数据的存在论意蕴与人类创构关联在一起。在与物能世界的关联中,人类涉及的最高层次相互作用是创造;在与作为信息数字编码的大数据世界关联中,人类涉及的最高层次相互作用则发展到无中生有的创生。所有这些都具有重要的存在论意蕴,都将在存在论层次给一些人提供全新的存在和发展条件,同时导致人群的进一步分化。

由于大数据不是技术层次的存在,以往任何技术都完全不能与之相比。大数据的发展将提供前所未有的更好——但其利用也相应要求更高的条件,从而空前凸显更具存在论意义的人群分化。“所有的技术变革都会产生有得有失的群体:赢家和输家。现有政治机构可能允许失败者阻止或减缓其采用,也可能允许成功者促进其采用。”^[16](P176-199)大数据越是发展到更高层次,人群的分化越是严重,数据鸿沟(Digital Divide)的形成和扩大,正是这种发展形势最引人注目的表现。

大数据的发展无疑会大大强化存在论层次人群的分化,所谓数字鸿沟就与此密切相关。早在21世纪之初的美国,数据鸿沟已经引起研究者的关注。“计算机和互联网的普及改变了美国社会的许多方面,从我们教育孩子的方式到我们与民选官员互动的方式。数字鸿沟这一概念的提出,旨在描述那些从这一新技术中获取优势的人与那些没有从中获取优势的人之间的沟壑。”^[17](P133-143)事实上,数字鸿沟远不只是技术鸿沟,而是具有存在论意义的发展事实。造成贫富差距还是表面上的表现,更根本的还在于,由于大数据越来越凸显的存在论地位,数据鸿沟造成的差别可以是存在论意义上的。由于信息数字编码的特殊地位,数字鸿沟在人类发展中的影响将会越来越深刻。它意味着,若能够更好地把握大数据,就可以获得大数据提供的升级存在基础;越是游离在大数据之外,则越可能相对而言反受其害。“如

果使用得当,数字技术使其成为可能的自由信息流日益丰富,但如果我们不知道如何接受它们……也许最重要的,不知道何时及为什么要关闭设备,或者让自己远离它,那么这些信息对我们个人来说会是不健康的,对企业来说是没有生产性的,对我们的社会来说是有害的。”^[18](P5)这也主要还只是涉及技术壁垒和传统伦理层次的问题,从理论上说,科学技术应当给整个人类带来普遍的福利,“任何人都不应被排除在获得和受益于这些对参与社会生活至关重要的资源之外”^[19](P79)。但在这个基本层面之上,还不可避免地存在有时候可能是巨大的差别,因为大数据的存在论意蕴涉及存在基础的变化。认为“‘数字鸿沟’是将人们分为两个群体的一种方式:有机会接入信息技术的人和没有机会接入信息技术的人”^[20](P483),还只是就技术壁垒而言,在更深层次,数据鸿沟不仅关系到数据领地,而且这些领地关乎两个不同世界的差别。涉及与物能意义上的领地相似的信息意义上的领地,无疑涉及大数据更深层次的存在论意蕴。在科学技术和社会发展过程中,有些方面的类似现象早就出现,但那相对于存在的基础,只是现象层次的补充,大数据则为此提供了存在论基础。

数字鸿沟所反映的,实质上是由于大数据为人的生存和发展提供了新的存在基础,从而使大数据利用的差别造成人群发展条件的不同。因此,从更深层次看,在人的发展过程中,数字鸿沟在基本的意义上意味着一种强大的张力,将大大促进人的相应发展。一方面,由于在根本上涉及以相互性(reciprocity)为基本特性的信息^①,在基本层面应当比以往任何科技发展带来的更是人类共同的福利;另一方面,科学技术成果给人类带来的共同福利越大,在人类中构成差别的能力也相应越大。在大数据的基础上,由于差别处于存在论层次,对人的存在和发展的影响就更为根本。这一点从大数据伦理问题的全新性质就可以典型地看到。

大数据的伦理问题不仅涉及新的技术创新,而且具有与一般技术创新不同的性质。大数据伦理不仅涉及对大数据的理解,而且需要对大数据有整体把握,因此具有存在论意义上的根本性。伦理问题越涉及根本,就越具有整体性;伦理问题越是整体性的,就越具有与基本规律和生存基础的密切关联,越具有存在论意义。这一层次的存在论意义,不仅由于人越来越以信息方式存在和发展,而且由于未来已来而人与人的存在和发展更根本地关联在一起。

在大数据发展的基础上,人类的存在和发展将是一个信息相互性不断增强的过程。人作为类群中的个体存在,在个体不断发展的基础上,大数据发展必定随着信息化越来越呈现整体化趋势。正因为如此,关于大数据基础上人类存在和发展的研究,就成了存在论研究新的重要课题。

参考文献

- [1] 王天恩. 大数据将把人类带进怎样的新世界? 解放日报, 2017-08-29.
- [2] Norbert Wiener. *Cybernetic, or Control and Communication in the Animal and the Machine*. Cambridge: The MIT Press, 1985.
- [3] 王天恩. 重新理解“发展”的信息文明“钥匙”. 中国社会科学, 2018, (6).
- [4] Bruno Latour. Tarde's Idea of Quantification//Matei Candea. *The Social After Gabriel Tarde: Debates and Assessments*. London: Routledge, 2010.
- [5] Danah Boyd, Kate Crawford. Six Provocations for Big Data// A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society. *SSRN Electronic Journal*, 2011, (9). SSRN, 2011-09-21[2021-02-18] <https://ssrn.com/abstract=1926431>.
- [6] Viktor Mayer-Schönberger, Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murry (Publishers), 2013.
- [7] 王天恩. 大数据中的因果关系及其哲学内涵. 中国社会科学, 2016, (5).
- [8] Robin K. Hill. What An Algorithm Is. *Philosophy & Technology*, 2016, 29(1).

① 相互性(reciprocity)是作为同一整体构成部分之间在整体过程机制中的相互依存、不可分割、彼此影响、共同行动的关系。

- [9] 詹姆斯·格雷克. 信息简史. 高博译. 北京:人民邮电出版社,2013.
- [10] Jules J. Berman. *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*. Waltham: Elsevier Inc., 2013.
- [11] 张轶群. 聚类成为大数据认知的突破口. 中国电子报,2015,(23).
- [12] 维克托·迈尔-舍恩伯格,肯尼思·库克耶. 大数据时代:生活、工作与思维的大变革. 盛杨燕、周涛译. 杭州:浙江人民出版社,2012.
- [13] 涂之沛. 大数据:正在到来的数据革命. 桂林:广西师范大学出版社,2013.
- [14] 吕乃基. 大数据与认识论. 中国软科学,2014,(9).
- [15] Colin Koopman. *How We Became Our Data: A Genealogy of the Informational Person*. Chicago: The University of Chicago Press, 2019.
- [16] Helen V. Milner. The Digital Divide: The Role of Political Institutions. *Technology Diffusion, Comparative Political Studies*, 2006, 39(2).
- [17] Kenneth R. Wilson, Jennifer S. Wallin, Christa Reiser. Social Stratification and the Digital Divide. *Social Science Computer Review*, 2003, 21(2).
- [18] Howard Rheingold. *Net Smart: How to Thrive Online*. Cambridge: The MIT Press, 2012.
- [19] Cees J. Hamelink. *The Ethics of Cyberspace*. London: SAGE Publications Ltd., 2000.
- [20] Michael J. Quinn. *Ethics for the Information Age*. Hoboken: Pearson Education, 2015.

The Ontological Implication of Big Data

Wang Tianen (Shanghai University)

Abstract As the product of the development of information digital coding, big data presents an information world different from the material world as a whole. As an informational existence different from the material existence, big data has important ontological implications. Big data means the upgrading of existence, as it increasingly exists not only in the form of data networks, but also by enabling things to exist online. The upgrade of existence has been brought by big data means the expansion from the material world to the data world, the network integration of informational and material existence, the promotion of higher dimensions of existence process and its grasp, and the integration of information levels of the existence and development of human and the outside world. Based on the big data form of information existence, we can not only get the big data view of the evolution of the meaning of ontology from ontology to existence, but also see from ontology to existentialism, from entity existentialism to relational existentialism, from the existence of the most basic matter to the existence of the highest level of human beings, and even from the "existence" to "nonexistence". Moreover, we can see the future dimension of existence, so as to see the deeper correlation between ontology and human existence and development on the basis of big data, and see the anthropological implication of ontology at a deeper level.

Key words big data; ontology; information encoding; existentialist implication

■ 收稿日期 2021-07-01

■ 作者简介 王天恩,哲学博士,上海大学哲学系教授、博士生导师,上海大学伟长学者;上海 200444。

■ 责任编辑 涂文迁