

人工智能时代的数据意识形态

——基于大数据对价值观和行为活动影响的思考

郑二利 王颖吉

摘要:在人人都无法脱离数据的大数据时代,大数据的影响不容小觑,它挑战了现代社会的既定秩序,也引发了新的社会问题。大数据给予未来世界无穷可能性,人们在欢呼大数据的美好前景时,也应该警惕新技术的威胁,理性看待大数据的影响。在大众传播领域,面对大数据技术,人是盲目和被动的,他的思想和行为会被诱导,会不经意间泄露自己的隐私,甚至侵犯他人的隐私。大数据技术把一切纳入算法中,大数据并不是完备的,算法也并不总是客观公正的,人们应该知道算法是否具有价值偏向、算法是谁的工具、它是为谁服务的。因此,人们需要了解大数据对人的价值观和行为活动的影响,了解算法渗透的意识形态性,警惕被数据和算法控制。说到底,数据既是社会商品,又是裹挟政治、经济、文化、资本等复杂权力关系的信息载体,它既关系到个体的虚拟身份认同,又关系到数据伦理的建构。

关键词:大数据;人工智能;数据意识形态;数据伦理

中图分类号:G206.3 **文献标识码:**A **文章编号:**2096-5443(2019)01-0072-08

大数据与人工智能革命的基石是数据,放眼当下世界,大到全球经济走势,小到个人网上社交行为,各类数据一同汇入庞大的数据浪潮之中,数据之大与数据传播之快速都是前所未有的。大数据技术产生之后,人们推测未来是人工智能制造神话的时代,由于人工智能善于学习,能利用算法提取数据,形成有效分析模型,因此它能模仿人类,甚至替代、赶超人类。利用人工智能医学诊断、自动驾驶、识别图像和语音等,都有非常好的前景。有学者预言,到超人工智能阶段时^①,人工智能会完胜人类。但是,人们在受益的同时,也应对其心存警惕。人类应该警惕自己对新技术的盲区,就在媒体欢呼大数据的美好前景时,大众在大数据面前往往显得盲目和被动。在大众传播领域中,大众的思想 and 行为常常被诱导,有时候不经意间泄露隐私,甚至侵犯他人隐私,大众对此却浑然不觉。因此我们应审视数据控制,并揭示其对人类观念和行为产生的影响。

其实,技术的历次革命,都会挑战传统秩序,继而影响人类的观念和行为。工业革命推动了启蒙理性,撼动了西方世界的神学秩序和王权秩序。大数据革命必然也会挑战现代社会的既定秩序,引发新的社会问题,如人与机器的关系问题、数据伦理问题等。生物科学、计算科学的发展,未来仿生生命极有可能威胁人类世界,科幻电影、科幻文学都表述过这种担忧,即在未来某个时间点,机器人侵犯人类世界,毁灭性灾难来临。总之,科技给予未来世界无穷的可能性,同时也可能引发种种危险性。而当下,我们正生活在一个任何人都无法摆脱数据的大数据时代,我们在满怀期待的同时也应心存警惕,理性看待大数据的影响。本研究着重分析在传播领域中,大数据对人的价值观和行为活动产生的影响,兹从以下几个方面展开论述。

^①人工智能发展划分为三个阶段,第一阶段弱人工智能阶段,第二阶段强人工智能阶段,第三阶段超人工智能阶段。目前人工智能发展还处在弱人工智能阶段。

一、大数据主义——一切纳入算法中

大数据即广义的数据,它是一切可以用作数据资料的统称,即可以被收集、测量、分析,能进行可视化处理的信息或知识,包含产生于自然和人类世界中的文字、图片、音频、视频等,以及各种物质材料,甚至宇宙粒子。人类使用数据的历史非常悠久,“已故的图灵奖得主吉姆·格雷在其《事务处理》一书中提到:6000年前,苏美尔人就使用了数据记录的方法,已知最早的数据是写在土块上的,上面记录着皇家税收、土地、谷物、牲畜、奴隶和黄金等情况。”^[1]传统数据是手工记录、整理的数据,主要服务于数字记录与统计。大数据是智能计算机运算、分析的数据。大数据一方面指数据生产规模扩大、种类多样化,海量数据自动生产;另一方面指人工智能可以学习和分析大数据,代替人进行预判。

如果说传统数据的主要职能是数字记录与统计,那么大数据的职能并不局限于此。大数据催生了人类对生命和世界的一种全新认识,即认为所有的生物都是算法生物,人和其他动物都是如此,也就是说,算法支配人和动物做出吃饭、睡觉、交友等行为。此外,当一切都是数据,甚至某些事物被概念化为数据时,那么一切都可以纳入算法的框架,而基于大数据的算法能描述包括自然、人类世界在内的一切。这样看来过去的金融、经济、政治等话题,其实都可以看成是基于数据管理和分析的算法问题。如以色列学者尤瓦尔·赫拉利谈到:“一般人看来,经济就是有农民种小麦,有工人做衣服,有顾客买面包和衣服。但在经济学家眼中,经济机制就是收集关于欲望和能力的的数据,再转化为决策”^[2]。在政治领域中,“自由市场资本主义和苏联式国家控制共产主义也不是意识形态、伦理教条或政治制度上的竞争,而根本是不同数据处理系统间的竞争。资本主义采取分散式处理,而苏联式共产主义则是集中式处理。”^[2]

可见,大数据建构了数据化世界理念,一切需通过数据呈现,大数据能呈现一切。同时,人类的自主判断会减少,而更依附于算法的分析和预判。比如在总统选举中,媒体利用大数据分析和描述总统和选民属性,预测选举结果;在体育赛事中,利用大数据分析球队的进球指数,预判冠军球队……在过去的旅行中,旅途充满了不确定性,人们需要按照地图提前规划行程,并留心公路上的指示牌,随时规划路线,防止走错路。有了数字地图,这一切都迎刃而解了。数字地图利用大数据解决了所有问题,它实时采集、分析用户数据,并上传共享,它能分分钟为用户规划行程。用户不必再担心旅途的不确定性了,因为数字地图提供的服务更及时,它比用户自己规划得更精准。

正因为大数据能消除不确定性,所以数据科学大行其道,它冲击了传统的政治、经济、文化等制度,并迫使其转型。数字经济已成为全球经济的重要组成部分,同时也成为传统行业转型的生长点,马云称其为“鼠标与水泥的结合”。文化领域同样受到数据科学的冲击,更倾向于依靠数据决策,如有的社交媒体利用大数据写稿。早在2015年9月10日,腾讯研发的机器人 Dreamwriter 写了首篇谈CPI增长的新闻稿,稿件内容是根据算法自动生成的。Dreamwriter 能根据指令在0.5秒内生成图文并茂的新闻稿,这是新闻记者办不到的。

在算法为王的时代,一切被纳入算法中,算法是否具有价值偏向、算法能否得出客观公正的结论、大数据是否一定是完备的数据等,这些都会左右算法的结果,进而影响人的观念和行为。算法是基于系统方法描述解决问题的指令,普通人几乎完全不了解算法的原理,而这恰恰就是盲区。今日头条的智能总工程师曾说,算法加加减减就能改变情绪,能把快乐的情绪通过算法加减变成悲伤的情绪。头条算法工程师的说法揭示了算法会被人干预的事实,看来它并不总是客观、公正的。或许正因此,2018年初今日头条宣布公开算法,今日头条的做法是为了树立客观公正的市场形象,告诉大众他们不会干预算法,今日头条的算法没有价值偏向。不过,让算法公开透明并不能保障客观、公正,因为数据体量大并不意味着数据是完备的。如果数据有问题,那么分析的结果也是有问题的,这也是需要警惕的。正如一位美国学者所说:“大数据不仅意味着有更多好的分析,也意味着有更多坏

的分析。要知道,即便是专家大牛也有掉链子的时候。如果一些不好的数据被心怀叵测的可疑人员添油加醋地利用,事情会变得更糟糕;不过,即便是动机纯洁的分析人员稍有不慎也会上当受骗。在这个充满数据的世界中,消费者得有一幅火眼金睛才行啊!”^[3]显然,大数据是非常复杂的,我们不能低估它的复杂性,尤其应该关注它是为谁服务的,它是谁的工具。在数据分析过程中,不仅仅数据在起作用,支撑数据的理论、定义数据的方式,都参与了数据管理与计算,都能干预算法。如果定义方式发生变化,数据理论发生变化,那么数据分析结果也会出现变化,这也应了头条工程师那句话,在算法加加减减之间,主导情绪变化。如搜索引擎中的头条、资讯平台中的头条,都有很强的导向性,百度就曾陷入为“莆田系”医院打广告的丑闻中。

当然,对每天裹挟着我们的各种社交媒体、资讯平台来说,大数据技术是其最得力的助手,在他们那里,庞大的客户数据就是资本,资本可以衍生权力和利益。对于普通大众来说,减少依附性,并对这一切保持理性变得非常重要。

二、数据管理与控制:隐形的权力争夺

在移动互联网普及的大背景下,数据量呈指数级增长,现实就是海量数据的堆积。显然,大数据的意义不只是体量大,关键还在于它的价值,即信息量大。香农在20世纪建构了信息论,用信息熵度量信息量的大小,解决了信息量无法测量的问题。“香农指出,信息量与不确定性有关系:假如我们需要搞清楚一件非常不确定的事,或是我们一无所知的事情,就需要了解大量的信息。信息的度量就等于不确定性的多少,这样香农就把熵和信息量联系起来。”^[4]信息熵代表不确定性,信息熵越大,不确定性越大,减少不确定性需要引入的信息量越大。

信息熵可以用来定义大数据时代的世界观^[4]。依据量子力学的观点,世界是测不准的,充满了不确定性。要降低世界的不确定性、减少信息熵,还需要引入更多的信息。大数据信息量大,更有利于消除不确定性。阿尔法狗打败了棋王李世石,就是因为它存储了足够多的围棋对弈数据,包括李世石下棋的数据,阿尔法狗有效利用大数据减少了棋局中的不确定性。引入更多信息消除不确定性,我们既需留意算法揭示的数据规律,也需警惕算法渗透的意识形态观念。因为算法是理论与数据的结合,它隐含着设计者的意图,设计者的意图支配计算的结果,这就导致结果出现价值偏向。就拿我们常见的搜索引擎来说,它们常常把商业广告放在首页中最醒目的位置,这种搜索结果明显是人为干预过的,也就是说,搜索引擎的算法体现了设计者的意图,即效益至上。

事实上,在大众传播领域中,数据信息并不能完全自由流通,也就是说数据信息虽然量大,但是并不完备。政治、经济等权力体系常常干涉数据信息流通,并试图占有、制造或阻碍信息。影视传媒公司操控数据的丑闻值得关注,有些制片方为了牟取票房暴利,在影片上映前刷票、锁场,通过伪造售票数据营造抢票信息,进而误导观众。制片方为了利益,不惜违规制造虚假数据,因为对他们来说,数据就是资本,控制数据就能控制观众。这里,数据控制的落差提出了全新的挑战,正如美国学者丹·席勒说:“大数据一方面,它会赋予那些靠信息获利的垄断资本更大的权力;另一方面,其他人却在毫不知情的情况下被剥夺掉一部分权力。这会造成更大的权力落差,预计将来会成为一个政治问题。”^[5]

当大数据关系到资本、权力时,数据争夺就变得更激烈了。也就是说,权力和财富的争夺有时表现为争夺和控制数据。对普通大众来说,更应该警惕被大数据隐含的资本和权力所操控。尼尔·波兹曼曾在《技术垄断》中谈到,“我们强大的意识形态工具就是我们的语言”^[6],因为人常常被深锁在自己语言的边界里,理解其对世界和人的描述。而浑然不觉那些边界外的人有着怎样的世界。同理,大数据时代,人们常常被深锁在数据结构中,依据数据思维、判断、抉择,甚至算法完全代替人做出选择,而人们对其中隐含的权力关系却浑然不觉。简言之,数据信息并不是客观、中立的,它有主动施为性,能导向人的价值判断和行为选择。

数据本身是中性的,而且庞大的数据堆积是没有意义的。为了实现数据的价值,必须把数据关联起来,并使其成为一种结构化的存在。在庞大的数据洪流中,有被关联的数据,也有不发生关联闲置的数据。管理和使用大数据需要数据挖掘技术,数据挖掘就是在一定范围内设定关键词,通过语法、词法、修辞分析,建立模型,处理数据。“在谷歌,至少有四成的工程师天天在处理数据。由于大数据的原始数据常常是没有固定格式、显得杂乱无章的,因此使用大数据的第一步是对数据的过滤和整理,去除与要解决的问题无关的维度,将与问题有关的数据内容进行格式化的整理,以便进一步使用。”^[4]不管是传媒业还是互联网、金融等行业,这些技术都已是标配。因为数据挖掘技术能帮助其从数据大杂烩中建立模型,分析数据,做出更有效的决策。显然,“这种借助数据进行决策活动的潜在机会不断涌现,其影响力涉及企业以及整个经济体的方方面面。正因此美国经济学家埃里克·布莱恩约弗森认为大数据将引发‘管理革命’。”^[7]

大众传媒领域的“管理革命”主要是针对用户的,用户被概念化为数据信息,这些数据信息被筛选、结构化处理后,用来预测用户可能带来多少利润,并方便管理者调整相应政策。针对用户的数据管理和分析,受市场意识形态支配,数据管理者考虑的是利润而非用户的利益。当用户信息被概念化为数据,用户被抽象化了,他的情绪、情感、尊严等这些无法数据化的指标多被忽略了。如果大众过多依附于大数据做出的决策,那么他们已经在不经意间被控制其中了。

最典型的现象是大众传媒中的数据推送和供给。围绕在人们周围的资讯平台、社交媒体常常会“主动”把信息推送给用户。看似这些数据的推送是依据算法为用户“量身定制”,但实际上主导这一切的是算法设计者的意图。当然,算法设计者的意图是非常复杂的,除了追求利润之外,还与意识形态和文化环境密切相关。社会的政治、文化、道德等因素也会介入其中,这些都间接影响数据推送,政经新闻推送便是如此。

数据屏蔽、数据过滤决定有些数据被关联起来成为有用的信息,有些数据被闲置或屏蔽成为“噪音”,信息屏蔽既是设立一种屏障,也是在设定价值导向。区分哪些数据是“噪音”,哪些数据是有用的信息,需要明确的判断标准。这里设立判断标准往往出于某种利益诉求,而非信息本身的质量和价值。资讯平台的数据推送,通常依据用户过往浏览记录筛选信息,对于吃货,那就给他美食类信息;对于购物狂,那就给他购物信息。显然,这种数据信息过滤只为更大概率占据用户视线,并不是出于为用户负责,而筛选的高质量信息。约翰·博斯利说过:“数据不仅仅是数字,它本质上源于人类的交互和社会协商,是一种社会商品。”^[8]约翰·博斯利只说了数据的一个维度,显然,数据既是社会商品,它还是裹挟着政治、文化等复杂社会权力关系的信息载体。在很多国家的总统选举中,在国际上的民族性事件中,在引起全球关注的各种政治丑闻中,都能见出这样的事实,即谁控制了数据信息,那他就控制了话语权。

总之,基于用户的数据分析剔除了多余的数据,保留有效的数据,在这个过程中,不是所有数据都被平等对待。那些指向性明确的数据信息,关乎用户既往经验的、符合用户喜好的信息,还有推送者有意推送的信息是有效的。显然,指向性明确的数据信息意在把用户控制在数据结构中,看似开放的虚拟社会被其设置了种种屏障和壁垒。越来越多的资讯、短视频“主动占据”用户的中心视点,试图引导和控制用户,这里既有国家意识形态的动因,也有资本逐利的动因,百度公司就多次被公众质疑为商户推广广告。毕竟,我们处在人人都无法摆脱数据的大数据时代,那些“主动占据”我们中心视点的信息,更可能被我们关注。反过来说,当数据成为资本,点击率意味着效益,针对用户的计算越来越精准,这也意味着用户可能被结构化、被固化,在某种程度上个性被消弭。

三、大数据镜像中的个体——身份分离

在人人都无法摆脱数据的大数据时代,人被概念化为数据,人的身份属性用数据描述。利用算法分析人的身份属性,是基于算法把人的身份数据进行结构化处理,这样做的原因是平台试图精准

定义用户。值得注意的是,算法并不能分析人的所有特性,也就是说,只有那些能纳入算法框架下的部分才能被描述,其他内容无法描述。一般来说,性别、年龄这些确定性的信息能用数据描述,除此之外,算法会着重分析行为偏好、个人趣味、甚至深层欲望等。但是,这些数据并不是完备的,因此人的虚拟身份与真实身份之间是分离的。这就需要我们思考数据镜像中人的身份问题,以及虚拟自我与他者之间的关系。镜像是拉康提出的范畴,他用镜像描述婴儿自我认同的过程,婴儿通过“镜中之像”区分自我与他者,并形成对自我形象的认知。在镜像阶段,婴儿的自我形象映射在对他的幻想上,通过镜像中他者的形象体验自己。但是,婴儿将自己映射到外部形象中,又会导致自我的疏离,这就是作为主体的自我与他者争夺的过程。数据镜像类似于拉康所说的镜像,不过二者性质不同。

数据镜像是虚拟镜像,是网络空间中那些标识用户身份的大数据形成的虚拟镜像。在网络空间中,几乎每个用户都有在社交平台、购物平台、支付平台等通用的身份。虚拟身份是由数据信息建构的,是用户在互联网上注册、浏览、购物、社交等产生的数据。标识用户身份的数据是元数据,元数据用来标识用户年龄、性别、社会地位、财富指数、文化指数、行为偏好、位置关系、个人趣味等信息。所谓元数据是关于数据的数据,是对事物进行命名、分类、结构化处理,用来管理数据的数据。美国学者克里斯汀·L·伯格曼说:“元数据是一种事物命名、数据表示和关系表示的手段……元数据是对信息资源进行描述、解释和定位等操作的结构化信息,其通过前述操作使信息资源更易检索、使用或管理”。^[1]元数据是管理数据的数据,用元数据描述用户身份属性,能把用户群结构化,并且精准化定位具体用户。这种操作终究是服务于平台检索、使用和管理用户,算法设计者通过算法引导、控制用户,而用户对这一切并不知情,用户常常不自觉地被平台推送的信息诱导就印证了这一点。可见,不管有意还是无意,用户似乎已经认同了数据镜像中的身份,最典型的表现是用户越来越依附于平台的定制和推送。在用户的身份数据中,算法侧重分析用户的个人趣味、行为偏好等信息,这些信息与用户的消费偏好关系密切。这与人的社会身份属性不同,人的社会身份关乎阶层、职业等与社会秩序相关的信息。相反,描述虚拟身份侧重关联人的消费偏好,而与其社会身份关系不大。说到底数据镜像强化了用户的消费偏好,而用户又受其诱导,常常会依此进行判断和选择。这一切导致人处于“信息茧房”中,难以主动进行理性分析,打破数据镜像。如电商用大数据“杀熟”,之所以有的电商能用大数据“杀熟”,就是因为算法把用户结构化为新老用户两类,并对新老用户实行不同的价格策略,吸引新用户采取的优惠政策会屏蔽老用户,这就出现同一商品老用户购买价格高,新用户购买价格低的情况。有意思的是,当老用户把既往信息全部清除之后,他的老用户身份就解除了,“数据镜像”被打破,他的虚拟身份又会被重新定义。可见,用户虚拟身份并不是同质的、稳定的,因为用户在资讯平台、购物平台、社交平台中所处的“结构”并不是一成不变的。

平台依据大数据精准把控用户,用户的数据量越大、数据越完备,运算误差越小。这里值得注意的是用户的数据信息并不会被平等对待,有些数据是活的数据,关联度高,有些数据关联度低,是被闲置的数据。关联度较高的数据,产生于社交平台、购物平台、支付平台、数字地图平台等,这些数据信息是“互信息”,能综合分析用户的消费和社交等情况,相反,其他与消费无关的个性化的信息并不被关注。

综上所述,虚拟身份与社会身份是分离的,社会身份是依据既往社会规则描述和确立的,与性别、年龄、职业、阶层、教育经历等这些因素相关,社会身份具有稳定性。平台并不在乎用户的真实社会身份,它更关注用户在虚拟空间的表现,特别是用户在虚拟空间中消费什么,关注什么,以及他的嗜好和偏好等。如有些平台的短视频推送,显然是利用算法分析了用户的偏好。因为其推送的短视频内容都是关联的,而且多与用户过去关注过的内容相关,也就是说,用户喜欢哪类短视频,平台就会推送大量同类视频给用户。当然,算法不仅计算内容,还计算视频时长、关联度等,多数两分钟左右的短视频以形式取胜,内容并无深度,形式却足够刺激视听感官。那些经过算法分析、筛选过的短视频能有效刺激用户神经系统分泌多巴胺,控制用户不停点击,迫使其“上瘾”。反过来说,如果用户

长期接受这些被推送的内容,他就会处于信息茧房中,会被“固化”。说到底,所有这一切都浸淫了消费社会的“毒”性。法国学者鲍德里亚认为,在消费社会中,消费不再是消极的消耗,而成为具有主导意义的模式,消费者也被消费社会依据消费占据的符号意义而重新定义。据鲍德里亚分析,在消费社会中,依据消费者占据商品的符号意义,消费者被归属于不同的阶层中。消费文化收编了启蒙运动以来个人幸福的含义,并使得幸福可以测量,换句话说在消费社会中,幸福的指数由购买力测得,所以鲍德里亚感慨,幸福有毒!而今,大数据技术被消费文化收编,消费者也被测量!

此外,数据镜像中的虚拟身份与人的真实自我也是分离的,因为数据镜像中的虚拟身份缺乏“总体性”和稳定性。大数据没有能力描述一个复杂的人,用户的身份数据信息都是不完备的。也就是说大数据只能计算纳入算法的数据信息,不能纳入算法的信息,大数据无能为力。人性是非常复杂的,人的情感、情绪、幻想、想象、联想、无意识等,这些非逻辑、非理性的部分无法用数据描述,不能纳入算法的框架中。算法是逻辑性的、理性的,不能纳入算法框架的信息就不能计算和管理。说到底,计算得清楚的东西未必重要,重要的东西未必计算得清楚。也就是说,人性中无法用数据描述的部分可能是最重要的,但却只能被放弃。而这些部分构成了真实的自我,也是自我区别于他者的核心所在。如果人不能警惕虚拟身份与真实自我是分离的,那么后果是很严重的。一方面,大数据规避了人性难测的一面,把复杂的人性抽象化为可以测量的指数,这沾染了消费文化的“毒性”;另一方面,年龄、性别、社会地位、财富指数、文化指数、行为偏好、位置关系、个人趣味等数据信息建构了用户的虚拟身份,但是用户却没有控制身份数据信息的权力。这些标识虚拟身份的数据很容易被别人控制和篡改,这无疑加剧了虚拟身份的不确定性,由此甚至引发伦理危机。最严重的莫过于用户数据泄露导致隐私泄露。对公众人物来说,身份失控后果更严重,尤其是公众人物一旦牵涉丑闻,他们的虚拟身份会完全失控,他者似乎拥有了书写、篡改、编造其身份数据的权利,而这一切却无需经过当事人知情或同意。大数据时代,公众人物更需小心翼翼维护自己的虚拟身份,维护其在大众心目中的“人设”,生怕其失控崩塌,毕竟,作为他者,大众对公众人物的认同也是通过数据镜像形成的。一切都是不稳定的,这似乎正应了马克·扎克伯格的说法,具有多于一重身份的身份是一种“缺乏总体性”的表现。

四、大数据的伦理悖论——他者的数据,我的隐私

今天,人们在互联网上几乎无处遁形,让人细思极恐的是我们的信息都在网上,而我们自己却无法保护自己的信息,因为我们几乎不知道这些信息的使用权和控制权在谁的手上。回想起来,最初的互联网是结构松散的虚拟世界,是一个没有人知道我是谁的空间,因为那时的用户无须认证,都是匿名上网,而且用户的虚拟身份与社会身份几乎没有关联。今天,不知道我是谁的时代已经终结了,大数据让我们在互联网上几乎无处遁形,同时,大数据时代常常发生个人数据信息被泄露,个人隐私被侵犯的情形。大数据引发的伦理危机频现,大数据时代,保护个人隐私、维护由于隐私泄露导致的名誉、权益受损,已成为重要的伦理难题。这里似乎存在悖论,一方面大数据时代需要个人生产和分享数据,提高数据运算的有效性;另一方面人需要保护自己的隐私,隐私数据不能共享。这看似形成了个人权益与公共利益之间难以弥合的鸿沟,实质隐含着数据资本争夺引起的伦理冲突和危机。

大数据时代,数据就是资本,拥有数据就拥有了资本。数据如此重要,以致有人用未来的石油比喻大数据。用户在社交媒体、网络平台、购物平台、数字地图等留下的注册、购物、浏览等信息都是数据,企业需要用这些数据计算用户,实现精准营销并获取效益。现在的高科技公司基本都是数据公司,苹果公司认为公司的未来就是发展人工智能,阿里的创始人马云也多次表示未来必然是大数据为基础的人工智能的世界。前谷歌研究院院长诺威格博士曾经谈到他在2001年加入谷歌,那时正是互联网泡沫破碎后,大家纷纷逃离互联网行业的时候,诺威格逆势加盟谷歌,因为他看中了“全世界的数据都在谷歌”^[4]。诺威格一语中的,揭示了大数据的价值所在。大数据需要大体量数据做基

础,数据体量越大,运算误差越小,运算效率越高。当数据能生产价值,制造数据、贩卖数据、控制数据、数据交易之类的暗箱操作就会频繁发生,这就导致用户数据信息被交易,用户隐私被泄露。

Facebook 因为数据泄露遭遇重大信任危机,而且面临巨额罚款。就在 Facebook 数据泄露危机还没有解除的时候,百度总裁李彦宏却说:“我想中国人可以更加开放,对隐私问题没那么敏感。如果他们愿意用隐私交换便捷性,很多情况下他们是愿意的,那我们就可以用数据做一些事情。”^①李彦宏的说辞引起了网民的吐槽和声讨。不过,从李彦宏的说法中,能看出百度公司的价值取向和对用户隐私不负责任的态度。支付宝也曾被质疑通过年度账单默认勾选、套取用户数据,其他各种社交平台、资讯平台套取用户数据、窃听用户信息以及泄露用户信息的行为也屡屡发生……李彦宏所说的情况现实中确实存在,部分用户为了享受价格优惠、获取资讯……会轻易把标识自己身份的数据信息留给商户,但是大多数用户其实并不知道自己的这些信息将被谁利用。当然也不乏这样的情况,有的医院、销售中心、中介公司、快递公司售卖并泄露用户的数据。猎豹移动安全专家李铁军在接受《中国经营报》采访时说:“你可以默认自己的信息已经被泄露,包括身份证、银行卡、手机号、邮箱等信息。”美国学者丹·席勒也说:“当自己启动洗衣机、打开电冰箱,冲澡或上床休息,都会留下信息。我们需要对此进行公开讨论,共同探讨决定我们的哪些信息可以被收集、被谁收集、作为何用,只有当这些问题弄清楚之后,才能解决大数据的问题。”^[5]大数据时代,个人隐私权正面临严峻的挑战,保护个人的隐私不被侵犯,需要个人提高警惕,防止数据信息被窃取,同时,也需要数据公司规范数据的使用权和控制权。当然,这种情况不是提高伦理意识可以完全解决的,还需要相应法律监管。

此外,网络暴力事件频发,也迫切呼唤数据伦理建构。普通大众也应提高数据伦理意识,确保自我能尊重他者的隐私。建构数据伦理,需要明确隐私数据的范围,知道哪些信息属于个人隐私,不能随意侵犯。此外,还需确保每个人的隐私数据都能享有平等地位,也就是说“数据隐私面前人人平等”。不过,在今天的网络世界中,似乎存在着一种吊诡的逻辑:他者的数据,我的隐私,而且历次网络暴力事件,都存在这样的伦理悖论。江歌事件、杨美芹事件等都是如此,网友们愤怒的“人肉”搜索,把当事人,甚至当事人的家人曝光在公众视线中。当事人不仅没有隐私权,而且还会遭受公众的舆论围剿,公众占据伦理的制高点对其进行伦理审判。更有甚者会找到当事人,并对其围追堵截,冒犯、骚扰当事人,让其无法工作和正常生活。在这样的事件中,网络审判的喧嚣掩盖了审判者对当事人的隐私伦理侵犯,也就是说,公众以反伦理的方式对他者进行伦理审判。在这类网络暴力事件中,公众显然置身于“他者的数据,我的隐私”这样的伦理悖论中。或许正因为公众没有隐私平等的意识,才会不自觉地侵犯他人的隐私,导致置身于网络暴力旋涡中的他者被剥夺了隐私权。而那些参与伦理审判的公众,集体作为“正义”的化身,共同参与了对他者的“窥视”和侵犯。

李彦宏说中国大多数用户愿意用隐私交换便利,这代表了商家对用户隐私的态度。在商家眼里,数据就是资本,它可以作为交易的对象,商家用“便利”换取用户的数据,用户把隐私“卖给”商家,只要二者达成“共识”,似乎不应该被谴责。其实,这里商家与用户的信息占位是不对等的。在用户看来,那些换取“便利”的零散信息,如年龄、性别、电话号码等,不能完全描述自己的身份,因此威胁较小,相对安全。但是,那些看似零散的数据信息,一旦相互关联起来,成为互信息,它的信息量就变大了。也就是说,个人在社交平台、购物平台、资讯平台、数字地图等留下的零散数据,一旦共享,信息互联,就能标识用户的私密身份了。算法的确有能力呈现这一切,小小购物车就能暴露用户最私密的那一面。从购物车信息中,能计算出消费者在饮食男女方面的偏好以及其经济状况、健康状况等。在商家眼里,能产生效益的都是商品,即便隐私数据也可以拿来交换,所以商人会回避数据伦理问题,而把一切定义为商品交易。显然,数据作为商品被“去伦理化”,是基于资本逐利的需求,事实上,隐私数据不能“去伦理化”。那些陷入泄露用户隐私丑闻的企业,很难再被用户信任,一次数据

^①来自 2018 年 3 月 26 日李彦宏在中国高层发展论坛上的讲话。

伦理危机可能给其带来灭顶之灾。当然,数据隐私泄露直接的受害者是用户,大数据时代,人人都无法摆脱数据,这关系到每个人的隐私安全,需要每个人理性对待。

参考文献:

- [1] 克里斯汀 L·伯格曼著.大数据、小数据、无数据:网络世界的学术.孟小峰,张祎,赵尔平译.北京:机械工业出版社,2017:5;53.
- [2] 尤瓦尔·赫拉利.未来简史:从智人到智神.林俊宏译.北京:中信出版集团,2017:334.
- [3] 冯启思.对“伪大数据”说不:走向大数据分析与解读的误区.曲玉斌译.北京:中国人民大学出版社,2015:5.
- [4] 吴军.智能时代:大数据与智能革命重新定义未来.北京:中信出版集团,2016:117;115;247;136.
- [5] 王建峰:告别信息崇拜解构数字资本主义:访美国伊利诺伊大学厄巴纳—香槟分校教授丹·席勒.中国社会科学报,2017-01-27.
- [6] 尼尔·波斯曼.技术垄断:文化向技术投降.何道宽译.北京:北京大学出版社,2007:72.
- [7] 史蒂夫·洛尔.大数据主义.胡小锐,朱胜超译.北京:中信出版集团,2015:103.
- [8] Hunter Whitney.洞悉数据:用可视化方法发掘数据真义.刘云涛译.北京:人民邮电出版社,2016:21.

Data Ideology in the AI Era ——About the Impact of Big Data on Values and Behaviors

Zheng Erli(Tianjin Polytechnic University)

Wang Yingji(Beijing Normal University)

Abstract: Everyone is unable to get rid of data in the age of big data, the influence of big data should not be ignored. Big data challenges the existing order in modern society and raises new social problems. In the future, big data can lead to a variety of possibilities. People need to deal with the impact of big data rationally, and when they are cheering the fine prospects of big data, they should be alert to the threat of new technology. In the field of mass communication, the public can only face big data technology blindly and passively, their thoughts and behaviors are often been induced, and the public inadvertently give away their privacy and even invade the privacy of others. Big data technology uses algorithms to measure everything. However, Big data is incomplete for Algorithm is not always objective and fair. Whether the algorithm has value bias, whom is the algorithm serving for, whose tools is algorithm, the public ought to understand them all. Thus, we need to be vigilant against data control and algorithm control, and rationally understand the impact of big data on human values and behaviors, and need to understand the ideology of algorithm. Obviously, data is not only a social commodity, but also the information carrier with political, economic, cultural, capital and other complex power relations. The above content is related not only to individual virtual identity, but also to the construction of data ethics.

Key Words: big data; AI Era; data ideology; data ethics

■收稿日期:2018-10-29

■作者单位:郑二利,天津工业大学人文学院;天津 300387

王颖吉,北京师范大学新闻传播学院;北京 100875

■责任编辑:刘金波