

# 计算社会科学研究： 范式转换与伦理问题

凌 昀 李 伦

**摘要：**大数据和数据化促成了社会科学研究范式的转变，催生了计算社会科学。计算社会科学具有超越传统社会科学研究局限性的优势，但也加剧了传统社会科学研究面临的伦理问题，同时在大数据背景下产生了新的伦理问题。为了确保大数据时代人的自由和权利，计算社会科学的研究应当更加重视社会科学研究伦理。

**关键词：**大数据；计算社会科学；科研伦理；社会科学研究范式

**基金项目：**国家社会科学基金重大项目“大数据环境下信息价值开发的伦理约束机制研究”（17ZDA023）；国家社会科学基金一般项目“开源运动的开放共享伦理研究”（17BZX022）

**中图分类号：**B089.3   **文献标识码：**A   **文章编号：**1003-854X(2020)09-0026-06

## 一、计算社会科学：数据化驱动研究范式的转变

如果把没有经过分析、没有产生意义或价值的现象或行为称为处于“自然状态”或是“混沌状态”的现象或行为，那么，把这种处于自然状态或混沌状态的现象和行为转化成可量化分析的数据并由此产生新价值的过程就可称为“数据化”（datafication）。我们无时无刻不在发生自然状态的行为，例如上下班的路径、其间使用的交通工具、顺路去什么地方采购等，但这些仅仅是没有经过数据化、处于“自然状态”的行为，或者即便我们将这些交通信息有意识地进行一定的分析和规划以提高自身的效率，这些信息也仅仅产生了小范围的影响或是少量价值。得益于“数据化”，我们能够清楚地认识到处于自然状态的行为或现象通过量化分析能够产生多大的价值，例如对一个城市人口上下班的路径、所使用的交通工具进行分析就有利于对城市的道路交通进行规划，而对人流量、周边住宅区域等问题进行分析对于建造购物综合体、农贸综合市场的选址而言都必不可少。由此可见，数据化能够将

个人的日常生活和行为进行量化，直接与行政管理和经济运转产生关联。

“数据化”是由肯尼斯·库克耶（Kenneth Neil Cukier）和维克托·迈尔—舍恩伯格（Victor Mayer-Schönberger）于2013年提出的概念，特指将人类行为和社会现象转化为计算机数据<sup>①</sup>。但是，数据化并不是近年来才出现的新趋势，数据化的现象和思维方法早已有之，近代自然科学就是数据化的先驱，即将自然现象或是经验现象还原为可量化分析的数据，再利用对数据的归纳分析得出自然的规律和一般理论。而将数据化从单纯的自然科学研究扩展到更广泛的社会问题研究同样也不是现在才产生的，早在19世纪，一些哲学及社会学学者就试图利用实验科学的方法进行社会问题研究。孔德曾提出“社会物理学”的概念，认为人是自然界的一部分，人类社会秩序是自然秩序的延伸，可以用物理学方法来研究人类社会。虽然，囿于当时的技术条件，将实验科学方法引入社会研究容易陷入机械论社会观的困境，但这无疑是将数据化引入社会研究的先驱，是计算社会科学诞生的直接思想源泉。

随着计算机技术和大数据技术的不断发展，现在的数据化能够更加深度、全面、准确地还原人类

行为和社会生活的轨迹，从而不仅为社会科学研究的数据化带来了新的生机，也能从研究方法和技术的角度回避机械论的困境。正如库克耶与舍恩伯格所说，（人类行为和社会现象的）数据化是新技术的发展促成的一种新趋势<sup>②</sup>。大数据技术和计算机技术为社会数据化提供了新的工具，使得计算社会科学应运而生。这个计算机科学与社会科学的新兴的交叉学科，通过新的计算技术和方法来进行社会模拟、建模、网络分析和媒体分析等，以研究社会 and 人类行为的关系和互动，其研究领域涵盖计算经济学、计算社会学、历史动力学、文化学以及对社交和传统媒体中内容的自动分析等等。

“社会物理学”一词目前仍在沿用，但其含义比最初提出时已有了很大的变化，指的是利用受物理所启发的数学工具来理解人类群体的行为<sup>③</sup>，在现代商业行为中，也指用大数据分析社会现象。因此，“社会物理学”与“计算社会科学”在当下的时代和技术背景下涵义基本等同。阿莱克斯·彭特兰对于社会物理学的研究就明显带有新技术的特征，他引入“想法流”（idea flow）的概念，认为社会物理学就是研究“想法流”的科学。在他看来，传统的物理学旨在了解能量的流动与运动之间的数学关系，而社会物理学旨在了解想法和信息的流动与人类行为之间的数学关系<sup>④</sup>。新想法的流动对于理解社会至关重要，新想法的传播和结合是行为转变和创新的推动力，而想法如何实现流动、传播和结合，都离不开数据化。

对比 19 世纪的社会物理学，计算社会科学所采用的大数据技术弥补了前者的理论缺陷；与传统社会科学相比，计算社会科学的不同之处则在于数据化驱动了研究范式的变革。大数据“正在引导一场计量方面的革命”<sup>⑤</sup>，对社会科学研究范式，尤其定量研究中的种种研究方法和手段，都产生了巨大影响。

从研究方法来看，计算社会科学对传统的调查分析、相关性分析、实验等方式都有不同程度的改进。在调查方面，计算社会科学不再局限于问卷或访谈等方式，而是通过大数据和算法来对庞大的数据进行处理和分析。大数据打破了空间的限制，使取样更为便捷和全面。更重要的是，计算社会科学拥有规避传统问卷调查一直面临的结果客观性偏差的技术手段——直接利用人类日常生活中相应行为的大数据，显然比本人填写的问卷更加客观。在实验方面，传统的实验方法包括实验室实验和实地实验，两者有其不同的优劣势。实验室实验在变量的

控制上更具优势，使相关性分析更为准确，但是选取的实验参与者可能并不那么具有代表性；实地实验更能反映真实情况，却不如实验室那样能准确把握变量。为此，有些研究者会在研究中将两者相结合，采取所谓的处于“实验室—实地”连续体上的实验。数据化则为这个二元的连续体添加了新的维度。马修·萨尔加尼克（Matthew J. Salganik）在传统的二元划分的基础上提出一种“模拟—数字”的实验模式。数字实验指的是利用数字基础设施招募参与者、随机分组、实施处理并测量结果的实验，而模拟实验则是在以上研究流程中完全不涉及数字基础设施的实验，许多实验可能并不处于这两极，而是处于其间的所谓“半数字”实验。“数字系统为实验—实地连续体上的所有实验都带来了新的可能性”<sup>⑥</sup>，传统的“实验室—实地”二元就变成了“（半）数字实验室—（半）数字实地”实验。

从研究对象来看，数据爆炸式增长极大地丰富了社会科学的研究样本，而机器学习、统计、社会网络分析和自然语言处理技术的不断升级，又强化了这一趋势<sup>⑦</sup>。同时，由于数据更加完整和准确，交互模型更加完善，新的交互关系也更容易被发现，催生了更多新兴交叉学科和领域，从而大大拓展了社会科学的研究对象。

然而，计算社会科学蓬勃发展的形势也不免引起一系列担忧。数据化往往会放大传统社会科学研究中所面临的伦理问题，或者使原本被极其关注的伦理问题不再受到重视，甚至完全被忽略。此外，数据化还催生了一些传统社会科学研究中不涉及或少有涉及的新伦理问题。下面我们将探讨三个在计算社会科学研究中可能或已出现的伦理问题，并透过这些显见的伦理风险探讨普遍数据化可能遭遇的更为严重的境况。

## 二、个人数据权利问题

计算社会科学研究不仅会采用源于特定研究的数据，而且会采用源于其他机构日常管理和营运的数据<sup>⑧</sup>，而数据来源的混合则容易产生侵犯个人数据权利等问题。

个人数据权利是人的基本权利在数据领域内的具体化。主张这一权利，使人能够控制自身数据，保障自由权、人格权和财产权等基本权利在数据领域内不受到侵犯。与之相对的，还存在机构数据权，即机构对于数据的采集和使用的权利。在计算社会科学研究中，存在着个人数据权和科研机构

数据权相对立的情况，其争端暴露了目前数据权属不明确的现状。传统社会科学研究中的数据收集目的相对单一，即科研，而计算社会科学研究中所使用的数据则不限于科研数据，还包括来自政府、金融机构、商业机构等用于管理和营运等采集目的的数据。这就产生了这样一种情况：也许我们愿意授权给政府、金融或商业等机构用于管理和营运目的使用我们的数据，从而获得这些机构的服务或商品，但并不希望自己的数据被用于社会科学研究。

当然，把管理和日常运行的数据用于研究目的并不是新现象，在临床医学领域早已有之。医学研究者会利用临床医疗数据进行医学科学研究，对新药、新技术使用进行后续跟踪分析等。传统社会科学研究也有利用已有的政府管理和企业营运数据进行研究的情况。但是，计算社会科学把其中的伦理问题放大了。一方面，社会数据可以涵盖更广泛的数据类型和使用领域；另一方面，医学研究也可能比社会科学研究更容易得到数据提供者的同意或默许同意。此外，数据的第三方授权使用，数据的外包处理，甚至是跨境流通等情况，使个人数据权利所面临的风险更为复杂多变，而个人数据权利的保障也更加困难。

一般来说，个人在机构面前都处于弱势，信息安全、数据泄露等会直接造成对个人基本权利的侵害。这种侵害可能是名誉性的，也可能会直接影响财产权和人的身心健康。尽管计算社会科学研究者们都声称研究数据经过了“匿名化”或者说是“脱敏”处理，但是这种“脱敏”效果可能并不尽如人意。由于人的信息被数个机构进行过收集，同一个人的相同或不同的数据可能同时存在于数个数据库中，即使各个数据库都没有显示敏感信息，但是将数据库进行关联分析就能够暴露出敏感信息，这就是所谓的“再识别攻击”。任何数据都有可能被再识别，并且所有的数据都可能是敏感的<sup>⑨</sup>。

这个问题看似陷入僵局，似乎只要存在数据的共享、传输，即使这种共享、传输是合规的，也同样无法避免个人数据权利受到侵害。实际上，我们在面对这种伦理风险时并非完全无能为力，创建并遵循一个动态的数据保护计划，不仅能够降低泄露数据的概率，而且能够在数据泄露发生后降低伤害。随着时间的推移，数据保护计划的具体项目，包括能够使用的加密形式都在改进<sup>⑩</sup>。从另一个角度来说，如果对个人数据权利的绝对保护只能基于数据完全保密的话，人们自然会去权衡数据化下的研究带来的利益和侵害。虽然我们会对风险有所担

忧，但大多数人不一定就此将自己与数据世界或这个信息时代完全隔离起来。

马修·萨尔加尼克从收益—风险二元维度分析了数据使用和个人数据权利的保护。在一个收益—风险体系中包括三种数据使用或数据保护的方式：非公开、保护性公开、无特定对象公开。非公开是一种极端的数据保护措施，完全不公开个人数据供他人使用，风险最小且收益最小；无特定对象公开是一种极端的数据共享使用措施，所有人都可以使用经过“脱敏”处理的数据（无论脱敏是否达到想要的保护效果），收益最大且风险最大；介于两者之间的就是保护性公开，也就是基于某种共识在达成共识的群体中进行数据的共享使用，这就是所谓的数据保护计划。虽然这种保护性公开仍然会遇到不少问题，或者说在不同的案例中会存在不同的问题，但这种根据不同的情境做出适应性调整的、动态的、协商的模式可能是应对目前个人数据权利和科研机构数据权利对峙局面的良策。社会科学的研究是为了在不同的情境下促进公共利益以及个人利益，如果我们完全禁锢个人数据，将对社会科学的发展乃至个体和社会的发展都造成阻碍。保护性公开措施为在数据共享中平衡个人数据权利和机构数据权利、个人权利和公共利益提出了一种可供选择的方案。

### 三、知情同意和隐私权问题

隐私权问题是大数据时代全社会各领域都在关注的伦理问题，主要表现在数据的使用和个人信息保护的争端上。把隐私权问题放入计算社会科学研究的情境下，还会涉及对知情同意的挑战。

如前所述，计算社会科学能够绕开传统的问卷调查，而将被研究者的相关行为数据化，直接获得研究所需的数据。传统社会科学中基于知情同意的研究伦理在计算社会科学中正遭受严重挑战，计算社会科学中更为准确而全面的数据可能正是以知情同意的缺失或隐私权的丧失作为代价而获得的。

当然，传统的社会科学研究也并不会在任何实验中都去征求所有相关方的知情同意，比如大多数研究就业歧视的案例中，会基于性别、种族等因素选取相应的被研究者，然后征求这些人的知情同意，但是少有研究会去征求雇主在这个实验过程中的知情同意，甚至美国法院也支持在实地调查中缺乏同意和使用欺骗来衡量歧视的行为<sup>⑪</sup>。如果我们在类似的对歧视问题的研究中以一种征得雇主同意



的方式进行实验设计和研究，那么被研究者反而会基于立场而做出偏颇的回答。同样，某些针对敏感问题或某些揭露性的暗访、卧底调查也不会征求知情同意。劳德·汉弗莱斯（Laud Humphreys）曾对圣路易斯城市公园公厕等公共场所里男性间普遍发生的、以即时满足和不用负责为特点的性行为进行社会学研究<sup>⑩</sup>。这项研究完全没有经过知情同意，就是为了避免这些在社会主流空间中颇具名望的人在面对如此敏感和私密的问题时，自我和社会自我的割裂而导致的言行不一。也就是说，知情同意原则与社会学的一些研究方法确实存在内在冲突。

而现在的情况可能更加糟糕，在计算社会科学研究中，完全不征求实验对象同意的情况可能会越来越多，即使是不需要暗访或是卧底调查的问题，知情同意也常常完全起不到作用。例如，我们可以时常见到一些大型的互联网公司下设的研究机构做出的各类报告，其中使用的数据涉及其不同领域的下属平台，包括社交、购物、阅读、金融服务等等。这些数据的使用仅仅是因为用户选择使用了这个平台，但用户用数据换取方便使用这个平台的机会，并不表明同意将自己的数据用于研究。再如，用户的数据常常在完全不知情的情况下被第一手采集者授权给第三方进行使用。除了知情同意被过于忽视，数据权属不明的现状也给知情同意带来了额外的障碍——知情同意中涉及了过多的相关方，而对象却不明确。

在隐私权保护方面，传统的社会科学研究就已经存在隐私权与研究结果准确性的内在矛盾，在计算社会科学研究中，隐私权的保护则变得越来越困难。一方面隐私权的定义向来模糊且具有争议，另一方面，以往的对于公共数据和私人数据的划分对目前的隐私权保护来说显然过于简单了，许多问题并不能简单地认为是公共数据和私人数据的冲突带来的，而是涉及了更加复杂的公共利益与各方利益的权衡。尼森鲍姆提出的“情境完整性”的分析方法给隐私权问题的讨论提供了一种新的视角<sup>⑪</sup>。

情境完整性针对的是特定情形下的信息流通规范，由三个要素构成：角色（主体、发送者、接受者）、属性（数据类型）和传输原则（数据流通限制）。在讨论某种数据传播和应用是否侵犯了隐私权时，需从这三方面进行考虑，而不是简单地认为数据应当完全共享或者共享即侵权。尼森鲍姆认为，隐私权既不是保密权，也不是控制权，而是适当流通个人信息权利<sup>⑫</sup>。尽管隐私权的界定还具有争议，随着时代和社会的变化，其在信息社会应

该会有不同于工业社会的内涵。但是，如何在以一定程度的欺骗、隐瞒获得更为客观的研究结果与保障隐私权、知情同意之间取得平衡，始终是需要解决的问题。

#### 四、算法偏倚问题

首先需要明确的是，社会科学研究中的偏倚现象是一直都存在的，即使在没有充分数据化的过去，当受访者了解了问卷或访谈所针对的问题以及研究的动机，他们也有可能基于本人的立场而做出刻意夸大或者忽视的回答，从而导致整个研究的结论被放大或是小觑。计算社会科学显然有弥补这一缺陷的优势，不过，来自技术层面的新的偏倚近年来已初露端倪，尤其在网络购物平台的数据分析应用上。这种来自算法的偏倚不仅会对计算社会科学研究的结果带来负面影响，更严重的是，一个有数据支撑的、似乎是客观准确的结论更容易成为政治、经济等领域公共政策制订的基础，而基于算法偏倚的公共政策将对社会产生不可估量的负面影响。“所有决策活动对经验与直觉的依赖将逐渐减少，而对数据与分析的倚重将与日俱增。”<sup>⑬</sup>这种对于数据能够带来客观结论从而更好地指导我们进行决策的推崇，让算法偏倚导致的后果更加容易被忽视。此外，传统的社会科学研究中存在的偏倚往往通过对变量、取样进行控制而容易得到校正，但是，算法黑箱让计算社会科学中的取样和因果关系变得不再那么透明，从而导致研究结果存在的偏倚难以得到校正。

计算社会科学是如何产生并延续整个社会的固有偏倚或偏见的？这是机器学习必然导致的结果。机器学习是对过去的、已经发生的数据的学习，如果过去的数据是因为某种偏倚或偏见而形成的，那么以此为基础的机器学习就是在延续这种已经存在的偏倚或偏见。如果计算社会科学仅仅是对这些由偏倚或偏见产生的数据进行分析并得出结论，那么由这样的结论指导的实践就会进一步将偏倚或偏见变得根深蒂固且难以应对。故而，比起自然科学或技术领域，机器学习在社会科学研究中的应用应当尤为谨慎。汉娜·瓦拉赫（Hanna Wallach）就从目标、模型以及数据三个方面，阐述了为什么机器学习在社会科学研究中尤为特殊。

从目标角度来看，计算机科学家和社会科学家有其完全不同的目标。计算机科学家注重预测，也就是用已经观察到的数据来推测丢失的或尚未观察

到的数据；而社会科学家的目标是解释，也就是为观察到的数据做出解释，再把这种解释与已有理论进行比较或提出新理论，所以社会科学家的研究往往建立在一个清晰的可实验的假设之上。从模型角度来看，计算机科学家和社会科学家的不同目标会导致不同的建模方式。基于预测目标的任务更注重预测的准确性而忽略因果性，因而模型结构可不受限制，一个需要大量数据进行训练的算法黑箱也是合理的；与之相对的解释目标则与因果关系密切相关，观察到的数据是对因果关系的佐证或是反对，因此模型必须是透明、可解释的。也就是说，预测模型通常被用来代替人的解释或推理过程，而解释模型是为人的推理提供信息或指导。从数据角度而言，同样基于目标不同，计算机科学家和社会科学家会选取不同类型的数据，计算机科学家会选取海量的、任何类型的数据用作任意预测目的，而社会科学家则基于具体问题来收集相对小规模的数据<sup>⑥</sup>。

由此可见，基于不同的目标及其衍生出的完全相反的建模方式，计算社会科学这一交叉学科可能比其他学科与计算机科学的交叉面临更多争议，这些争议主要来自于伦理方面。也许这些伦理问题对于社会科学家而言并不陌生，或者说，伦理考量是社会科学研究工作中必不可少的环节，但是对于计算机科学家来说，这可能是容易忽略的问题，是他们曾经的工作中不必然会涉及到的。因此像计算社会科学这样的跨学科研究在使用计算机方法的时候，必须谨慎考量伦理道德风险。

总而言之，一方面，就像帕梅拉·麦考达克（Pamela McCorduck）所指出的，对于女性和少数群体而言，“我更倾向于一台公正的电脑”<sup>⑦</sup>。她认为，在某些环境中，我们会更加倾向于自动法官和警察，而不需要人为的裁决<sup>⑧</sup>。算法和机器学习对于社会科学研究消除偏倚或偏见是必要且有其优越性的，它把人的主观偏见从决策过程中剥离了出来。另一方面，就像汉娜·瓦拉赫所说的，计算社会科学不是计算机科学和社会数据的简单叠加，机器学习对于计算社会科学而言并不是一劳永逸的方法，如果我们想要以一种负责任且合乎伦理的方式利用机器学习来理解社会现象，我们仍然需要人类社会学者工作<sup>⑨</sup>。

## 五、数据巨机器问题

在评价 19 世纪的社会物理学时，许多人认为，当时的社会物理学学者把社会视为一部机器，人的

主观能动性和人与社会的交互被忽视，社会的变迁仅仅是“外力”的结果，故而造成了当时社会物理学理论的缺陷。随着相关技术和学科水平的提高，数据化越来越能够展示出人与人之间的差异性以及人与社会的相互影响，但是在这种理论缺陷得到弥补的同时，一种“数据巨机器”的可能性又展现在了我们眼前<sup>⑩</sup>。

如前所述，机器学习可能会延续已有的偏倚或偏见，而当这种带有偏倚或偏见的理论被用于实践和决策时，就有可能固化偏见，并使其变得更难以动摇。由此可以发现，在计算社会科学中，虽然人与社会的交互作用能够被发现，但是研究本身作为一种实践活动，其与社会的交互却可能被忽视。这种由研究与社会的交互带来的互相影响、互相支撑从而将谬误忽视掉或合理化的现象，就是所谓的“操演性”（performativity）干扰。

所谓“操演性（或述行性）”，是指当一个理论以使世界和理论更加符合为目的地改变着世界的时候，我们就说这个理论具有操演性<sup>⑪</sup>。也就是说，计算社会科学研究不仅仅在解释社会现象，同时也在改造社会使其发展与研究本身相符合。这种“操演性”在计算社会科学中的影响会比在传统社会科学中的影响更大且难以发觉。例如亚马逊 AI 招聘。亚马逊的电脑模型通过对过去 10 年提交给该公司的简历进行学习，由此得出男性求职者适合电子商务行业的结论，以此来审查应聘者，并向男性推送更多的此类应聘广告。这种通过机器学习得出的“男性比女性更适合电子商务行业”的结论就可能会通过招聘这一实践行为的结果而得到固化，使这个行业真的成为男性主宰的行业。

同理，计算社会科学的发展对于“数据巨机器”的形成也可能会有这种操演性影响。正如前文所说，与传统社会科学基于知情同意并以科研为目的来收集数据的研究方式不同，计算社会科学可以直接利用政府或是其他商业机构用于日常管理以及提供服务的现成数据来进行研究，那么整个世界都将可能成为一个泛化的实验室，样本和总体再无差别，甚至传统的“实验室—实地”二分的实验模式在社会科学领域都可能会被消解，关于人和社会的所有数据都可以成为计算社会科学的现成样本，从而形成一个基于社会科学研究的“数据巨机器”。

此外，计算社会科学通过将问题进行“理论概念”使之能够进行定量分析，再将其与某些被数据化的行为或是现象联系起来，从而使这些经过数据化的数据能够作为被研究的社会问题的证据。例

如,信用值就是将信用这一抽象要素通过构念变为一系列能够被量化的指标,比如是否按时偿还贷款、是否有违法犯罪记录、租赁是否合规等,然后依此将信用通过数据表现出来。这就导致了这样的情况,如果我们缺乏关于被理论构念的指标的数据,就可能难以构建起我们的数字身份,从而被计算社会科学研究边缘化,也无法获得后续的应用或服务。反之,如果想要在社会问题研究中占有一席之地,就必须将自己以某种方式数据化。也就是说,计算社会科学的发展要求全面的数据化以及数据的自由流通,而这就是数据主义的追求。

数据化从本体论和认识论的角度全面影响人们对世界的认识和把握,催生了数据主义的思潮。数据主义最初只是对大数据、数据化的哲学表达,是大数据时代所产生的一种哲学或是理念,但是随着数据化对人的日常生活决策和社会运行的影响不断加深,其正在转变为一种崇尚数据自由至上的思潮。正如赫拉利所说,“数据主义一开始也是一个中立的科学理论,但正逐渐成为要判别是非的宗教”<sup>②</sup>。

一方面,在社会科学研究中,有着数据支撑的、客观准确的结论受到推崇,于是为了获得这样的结论,或者为了促进学科发展和公共利益,个人的权利和自由似乎就变得无足轻重了。另一方面,受自身操演性影响,计算社会科学可能会以一种数据主义的方式改造世界,直接加速数据巨机器的形成,导致人成为只是可以被任意使用、流通的数据。也就是说,在数据化影响计算社会科学的同时,计算社会科学的研究方法、研究成果的应用反过来巩固了数据主义的地位,并可能生成数据巨机器。数据巨机器犹如“楚门的世界”:在楚门的世界,除了楚门,人人都是演员;在数据巨机器里,人人都是楚门,人人都是演员,无人是自己<sup>③</sup>。目前,大数据能够影响人的自主选择已成现实,例如各平台的大数据杀熟,剑桥分析公司对多国竞选的操纵,亚马逊AI招聘中的性别、种族歧视,等等。为了避免数据对人的自主行为的进一步操控从而形成数据巨机器,避免人被数据技术所控制而成为技术所奴役的对象,沦为这个数据巨机器中的螺丝钉,计算社会科学应当重视自身可能涉及的伦理问题和伦理风险,重视人在数据领域的权利和自由。作为解释人类社会的学科,计算社会科学应当从一开始就自觉地担负起这一责任。

#### 注释:

① Kenneth Cukier and Viktor Mayer-Schoenberger, *The Rise of Big Data*, *Foreign Affairs*, 2013, 92(3), pp.28-40.

② 维克托·迈尔-舍恩伯格、肯尼斯·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,浙江人民出版社2013年版,第123—124页。

③ Gerard George, Martine R. Haas, Alex Pentland, *From the Editors Big Data and Management*, *Academy of Management Journal*, 2014, 57(2), pp.321-326.

④ 阿莱克斯·彭特兰:《智慧社会:大数据与社会物理学》,汪小帆等译,浙江人民出版社2015年版,第8页。

⑤⑬ 史蒂夫·洛尔:《大数据主义》,胡小锐、朱胜超译,中信出版集团2015年版,第6、9页。

⑥⑨⑩⑪⑫ 马修·萨尔加尼克:《计算社会学》,赵红梅、赵婷译,中信出版集团2019年版,第212、426、427、414、45页。

⑦ W. Mason, J. Vaughan, H. Wallach, *Computational Social Science and Social Computing*, *Machine Learning*, 2014, 95(3), pp.257-260.

⑧ 尽管传统的社会科学研究也可能采用其他机构产生的日常数据,但无论范围和规模,都与计算社会科学研究不可同日而语。

⑫ 这个案例已成为臭名昭著的社会科学研究伦理案例。

⑬⑭ Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, *Stanford Law Book*, 2010, pp.67-89.

⑮⑯ H. Wallach, *Computational Social Science ≠ Computer Science + Social Data*, *Communications of the ACM*, 2018, 61(3), pp.42-44.

⑰⑱ Pamela McCorduck, *Cli Cfe, Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*, A. K. Peters/CRC Press, 2004, p.356, pp.374-376.

⑳㉑ 李伦:《数据巨机器的“意识形态”——数据主义与基于权利的数据伦理》,《探索与争鸣》2018年第5期。

㉒ 尤瓦尔·赫拉利:《未来简史》,林俊宏译,中信出版集团2017年版,第346页。

**作者简介:** 凌昀,大连理工大学大数据与人工智能伦理法律与社会研究中心、人文与社会科学学部博士研究生,辽宁大连,116023;李伦,大连理工大学大数据与人工智能伦理法律与社会研究中心、人文与社会科学学部教授,辽宁大连,116023。

(责任编辑 胡静)