当代社会科学研究中的解释非正义

Hermeneutical Injustice in Contemporary Social Science Research

白惠仁 /BAI Huiren　　郭卓睿 /GUO Zhuorui

（浙江大学哲学学院，浙江杭州，310058）
(School of Philosophy, Zhejiang University, Hangzhou, Zhejiang, 310058)

摘　要：基于大数据的社会科学研究中隐藏潜在的偏见与不平等问题。然而，大数据的客观性神话使得当代社会科学产生了“大数据傲慢”的研究倾向，即认为大数据可以代替传统社会科学分析。鉴于社会科学对当代社会集体解释资源的强大塑造作用，忽视大数据偏见会使未被纳入大数据集的弱势群体处于解释上的边缘化地位，从而遭受弗里克所称的解释非正义。为了应对解释非正义，社会科学共同体需要保持认知谦逊，即公开承认和反思大数据方法的局限性，并积极寻求其他认知资源。

关键词：大数据傲慢　解释非正义　认知谦逊

Abstract: Data-intensive social science research conceals biases and inequalities. However, the misplaced belief in the objectivity of big data has fostered what is termed “big data hubris”, the assumption that digital methods can fully replace traditional approaches in social science. Given the crucial role social science plays in shaping society’s collective hermeneutic resources, neglecting the biases inherent in big data risks marginalizing underrepresented groups, leaving them excluded from datasets and thereby subject to what Miranda Fricker describes as hermeneutical injustice. Addressing this injustice requires the cultivation of epistemic humility within the social science community. This entails critically examining the methodological limitations of big data, openly acknowledging these shortcomings, and actively pursuing complementary research strategies to ensure a more inclusive and equitable framework for knowledge production.

Key Words: Big data hubris; Hermeneutical injustice; Epistemic humility

中图分类号：C91; N031　DOI: 10.15994/j.1000–0763.2025.07.002　CSTR: 32281.14.jdn.2025.07.002

当代科学具有显著的“数据密集型”特征，而本就具有数据需求和实证传统的社会科学也在“万物皆数据”的数字化浪潮下走向基于大数据的“第四研究范式”。[1]这一新兴研究范式以计算社会科学（Computational Social Science）的迅速发展为标志，计算社会科学旨在利用日益增长的数据量和计算能力揭示传统方法难以捕捉的社会现象。2022 年，欧盟联合研究中心（Joint Research Centre, JRC）启动了“为政策服务的计算社会科学”（Computational Social Science for Policy, CSS4P）项目，并发布了关于计算社会科学政策需求报告（Mapping the Demand Side of Computational Social Science for Policy）。[2]该报告试图从需求角度提出计算社会科学的发展方向，不仅需要评估科学如何为政策提供支持，还要考虑政策制

基金项目：国家社会科学基金青年项目“当代科学中的认知非正义问题研究”（项目编号：22CZX026）。
收稿日期：2024 年 10 月 12 日
作者简介：白惠仁（1988–）男，内蒙古巴彦淖尔人，浙江大学哲学学院“百人计划”研究员，研究方向为知识正义、科技伦理。Email: baihuiren@zju.edu.cn
郭卓睿（2001–）女，山西运城人，浙江大学哲学学院硕士研究生，研究方向为科技伦理。Email: guozhuorui@zju.edu.cn

定者如何从一开始就参与CSS4P议程。这是因为社会科学研究服务于社会治理时，其研究结论会对人类社会产生难以预见的伦理影响。其中，偏见问题是计算社会科学研究面临的重要挑战，同时也与社会正义问题密切相关。由于社会人口统计变量影响数据生成，现有的社会不平等往往会通过数据集嵌入到社会科学研究中。当有偏见的社会科学结论转化为社会政策时，可能会固化甚至放大社会不平等。尽管偏见问题长期存在于社会科学研究中，但是当前社会科学研究者对技术的过度依赖可能会使他们更容易忽视潜存的偏见和不平等问题。

鉴于当代社会科学研究在个体生活和国家治理中发挥日益重要的认知作用，学界开始从多角度探讨其伦理问题。从技术角度出发，《计算社会科学手册》（*Handbook of Computational Social Science*）将伦理问题作为重要专题，霍林斯黑德（William Hollingshead）等人提出，计算社会科学所面临的伦理挑战主要涉及数据代表性问题、数据存档中的隐私问题、数据聚合与分析技术，以及数据链接技术相关的伦理问题；[3] 从应用角度出发，《为政策服务的计算社会科学手册》（*Handbook of Computational Social Science for Policy*）探讨了计算社会科学作为决策工具的优势与社会风险，既涵盖了数据正义、公共福利等一般问题，又讨论了计算社会科学在就业分析、社交媒体及经济发展中的具体应用问题；[4] 从范式转换角度出发，凌昀等人讨论了计算社会科学中个人数据权、数据隐私以及算法偏见等关键问题。[5] 然而，对于计算社会科学伦理问题的讨论仍存在不足：第一，社会科学研究伦理与大数据伦理缺乏交叉点①，例如，在处理数据隐私问题时，若未考虑到研究方法可能造成的影响，研究结果的有效性就会受到质疑；[6]，[7] 第二，现有研究几乎未涉及数据技术对人们认知实践的影响，既包括研究者对数据技术的认知依赖，也包括被研究对象可能遭受的认知伤害；第三，当前讨论主要延续了大数据伦理的进路，缺乏分析社会科学大数据伦理的理论框架②。

本文引入弗里克（Miranda Fricker）的解释非正义（hermeneutical injustice）概念，考察当代社会科学研究中被排除在大数据之外的弱势群体所遭受的认知与伦理层面的双重伤害。[8] 解释非正义是指社会集体解释资源的系统性缺乏使得某些人在解释其社会经历时处于不利处境，是认知非正义（epistemic injustice）的主要形式之一。（[8]，p.1）在社会科学研究结果转化为政策的过程中，社会科学研究通过影响社会制度与规则塑造集体解释资源。因此，当社会科学研究忽视弱势群体的观点或行为痕迹时，这些群体将面临“解释鸿沟”，即缺乏必要的解释资源理解并表达自身重要经验，导致其社会处境进一步恶化。由此，本文将通过以下框架讨论当代社会科学研究中的解释非正义问题：第一部分分析当代社会科学研究忽视大数据偏见的原因，即大数据的显著优势使许多学者在研究中对其过度依赖，产生了大数据傲慢（big data hubris）；第二部分提出当代社会科学研究中源于大数据傲慢的解释非正义，社会科学研究成果是社会解释资源的重要来源，这使得被忽视群体难以自我理解或向他人解释自身的被压迫经验，并可能遭受更普遍的社会非正义；第三部分则针对社会科学中的解释非正义提出基于研究共同体的认知谦逊（epistemic humility）作为应对方案，认知谦逊既要求研究共同体的反思批判意识，又需要其公开承认研究局限性以寻求其他资源弥补缺陷。

①拉泽尔（David Lazer）等人认为计算社会科学领域中的科研伦理、教学方法以及数据基础设施仍处于萌芽状态，还有许多制度结构问题亟待解决；斯托迈尔（Stommel Wyke）和赖克（Lynn de Rijk）的研究表明，64%使用大数据的社会科学研究没有讨论伦理问题。

②当代科学实践主要包括科学决策、科学研究、成果应用三个阶段，不同阶段伦理问题的具有不同特征。目前的社会科学伦理研究并没有区分科学研究的伦理与成果应用的伦理，这实际上体现出今天应用型科学研究的主要特征，即科学研究的价值、规范越来越多义和模糊。基于社会科学研究与社会之间的深刻互动，本文也并不严格区分研究伦理与成果应用伦理，而是以社会科学研究共同体为中心。

一、当代社会科学中的大数据傲慢

认知傲慢（epistemic hubris）指个体对复杂问题的过度自信，表现为缺乏专业知识的个体通过有限的研究误以为自己可以替代领域专家。[9]同样，大数据傲慢描述了一种研究者对数据科学技术的过度自信，误以为数据科学技术能提供更高形式的智能和知识，从而替代传统社科研究方法和专家的作用。[10]计算社会科学家拉泽尔提出，大数据傲慢通常是一种隐含的假设，即大数据完全替代了传统数据收集和分析，而非补充。[11]梅塞利（Lisa Messeri）与克罗克特（Molly Crockett）在《自然》（*Nature*）杂志上进一步指出，尽管AI工具承诺通过超越人类认知限制提升科学研究的生产力与客观性，但过度依赖可能带来理解错觉（illusion of understanding），即对知识深度和广度的误判。这种对大数据的过度信任和依赖还会导致科学的单一文化（scientific monocultures），即某一种知识生产方法被广泛采用并排斥其他替代方法，从而使知识生产缺乏多样性。[12]乌普里查德（Emma Uprichard）也认为，仅关注主导模式的大数据研究忽视了社会分裂、压迫和边缘化声音，导致“方法论种族灭绝”（methodological genocide）。[13]

部分社科研究者过度夸大大数据的客观性主要有两方面原因。一是，社会科学对“客观性”的长期追求使研究者对数据本身具有高度敏感性，他们倾向于使用标准化技术手段获取数据。从传统的田野调查、问卷研究到今天的大数据技术，这些标准化技术都有其正确应用的数据处理协议，以确保数据的完备性和有效性。例如，问卷调查的处理协议不仅规定了问卷设计及实施的正确方法，还包括预防数据收集过程中可能出现误差的措施。这些技术及其产生的数据都被认为是客观中立的，可以形成一系列可靠事实，从而服务于理论化的目标以及对理论的验证。[14]二是，大数据的确为社会科学提供了不可忽视的认知价值，其以高速、客观的方式使研究者在信息获取、问题挖掘和研究方法方面取得了显著进展。在社会科学第四研究范式下，研究主体由单纯的科研人员转变为大数据系统与人组成的共同体，甚至可以自主完成数据获取、建模和分析预测的任务。[15]然而，部分学者已经意识到人让位于计算系统的风险，提醒社会科学研究者如果过分沉迷于机器预测和操控而忽视了社会理论的价值因素，就可能会面临被彻底物化的风险。[16]这意味着，尽管大数据技术在一定程度上提高了研究效率和客观性，但也可能导致社会科学研究陷入过于简化和机械化的境地，从而忽视社会现象的复杂性和多样性。

从数据角度看，尽管大规模数据集减少了传统小数据采集的限制，但其客观性也不应被过度夸大。《大数据时代》一书指出，大数据的一大特征是实现从样本到总体的转变。[17]据此，一些社科研究者认为，大数据是社会信息的集合，能够突破传统社会科学有限数据的限制。[18]然而，数字地理学、[19]后殖民主义[20]以及社会学[21]从不同角度对大规模数据的客观中立性提出质疑。《中国数字治理发展报告（2023）》显示：全球有37%的人口从未接触过互联网，而在中国约有4亿非网民面临互联网基础设施不足、文化教育水平较低、缺乏数字技术使用技能等社会公平问题。[22]罗杰森（Simon Rogerson）等人指出，因经济、地域和教育等因素，“数字弃儿”在数字化时代难以获得完整意义上的存在。[23]这些弱势群体在数字社会中被明显边缘化，面临数字排斥，即在使用信息和通信技术方面存在着不平等的机会和能力。因此，数字不平等是更大的社会不平等的一部分，不仅表现为个人层面的边缘化体验，还与贫困、教育、性别歧视和地位差异等社会因素相互作用、相互强化。

社交媒体数据的易得性使其成为社会科学大数据的主要来源之一。近年来，关于社交平台大数据的研究呈现爆发式增长，研究者从最初盲目相信社交平台大数据具有“总体”特征到关注其用户代表性问题。基于美国成年人社交网站使用的全国样本调查，哈吉塔（Eszter Hargittai）利用实证方法揭示了社交平台大数

据的潜在偏见。[24]不同平台用户具有不同的社会人口特征，比如，相比于Twitter用户，使用Facebook的用户往往具有更高的教育水平、收入和社会地位。此外，社交媒体平台的标签功能以及平台的用户分组抽样问题也使得社科研究者所收集来的数据难以完全代表特定平台的用户和内容。主导群体的观点和行为更易被纳入社交媒体大数据中进行研究，而弱势群体的声音常被忽视，这导致政府和企业依据偏见数据进行决策时，忽略弱势群体的真实需求。因此，过度依赖大数据进行研究可能会使一个更加完整和真实的社会消失于社会科学的研究视野之中，这种依赖会限制对社会多样性的理解，并加剧政策制定和社会服务的不公正。

从研究角度看，尽管偏见问题长期存在于社会科学，但是数字化方法可能使研究更易受到破坏性偏见（disruptive bias）的影响。在社会科学研究中，拥有不同社会地位的人对他们所处的世界有着不同的理解，而研究问题的界定、方法设计以及数据收集分析往往取决于社会科学研究者看待世界的角度。[25]社会中许多现象对处于优势地位的研究者而言是“不可见”的，这使得他们难以将这些现象纳入重要的研究议题。在后殖民主义对现代社会科学知识的反思中，主流思想流派被称为“北方理论”，而产生于后殖民时代的澳大利亚、非洲等地的理论则被称为“南方理论”。[26]这一区分的核心在于批判现代社会科学在普遍性知识的表象之下，隐含着宗主国社会的观点和视角。女性主义和批判理论同样认为，由于性别、种族、阶级等特定社会因素，社会科学很难认识到关于人类社会的一些至关重要的部分。[25]沙利夫扎德（Rahman Sharifzadeh）进一步指出，数据科学技术更易引发社科研究中的破坏性偏见。与知识政治中不可避免的本构性偏见（constructive bias）不同，破坏性偏见对研究的内在逻辑构成了更深层次的威胁。[27]大数据知识生产依赖于科研平台和科技巨头的海量数据，要求技术与资本的深度结合。在这一过程中，大型科技公司优先服务经济目标，而非推动社会科学研究的认知目标，这种机制可能导致社会科学研究偏离其初衷，忽视对真实社会现象的理解。

综上，由于大数据的客观性神话，许多社会科学研究者产生了大数据傲慢的研究倾向，并在研究中过度依赖大数据技术。然而，大数据不可避免地带有预设和标签，这就意味着其中包含着难以察觉的偏见，当这些偏见以某种形式存在于大数据之中时，传统社会科学研究所具备的敏感性和批判性就显得尤为重要。如果研究者出于对大数据客观性的盲目崇拜而忽略这些潜在问题，他们就会降低对数据背后深层次偏见的敏感性，从而使偏见在社会科学研究中变得更加根深蒂固。大数据在社会科学中逐渐构成多样化的知识生产要素，同时可能形成新的认知权威，导致创新受限于既定范式，使研究者从积极的世界构建者转变为被动的预测对象，最终损害社会科学自身。

二、源于大数据傲慢的解释非正义

“认知非正义”是弗里克融合认识论与伦理学所提出的一种认知层面的非正义概念，即通过认知实践或认知制度非正义地对待某些特定知者，从而引起认知障碍，导致认知与伦理层面的双重伤害。其中，解释非正义是弗里克所刻画的认知非正义的主要形式之一，由社会解释资源的不足直接引起。“我将这种认知非正义解释为源于集体解释资源的缺乏——存在于我们共享的社会解释工具中——这种缺乏导致的认知劣势对不同社会群体造成不平等的冲击并非偶然。”（[8]，p.6）进一步，解释资源长期、结构性的缺乏引起了解释边缘化（hermeneutical marginalized），“不平等的劣势源于最弱势群体成员在某种程度上遭受解释边缘化——也就是说她们不能平等地参与生产社会意义的活动。”（[8]，p.6）社会弱势群体往往由于经济条件或是种族、阶层等身份原因无法平等参与塑造集体解释资源的认知实践。例如，在第二波女性主义时期，新闻、政治、法律等领域爆发了很多开创性的女性主义活动，其中的重要原因就是当时女性相较男性处于弱势地位，阻碍了她

们平等参与新闻、政治等生产社会解释资源的活动。社会意义通过认知实践产生，如果一个群体的成员无法充分参与这种实践，那么这个群体就被认为处于解释边缘化地位。

弗里克将解释非正义归结为一种纯粹的结构性因素，即对话双方都面临解释资源不足，但说者因在社会中总体处于劣势而无法清楚地理解或表达对自身十分重要的经验。在“产后抑郁症”概念出现之前，女性很难向他人表达产后由于体内激素紊乱而导致的一系列身心异常现象，甚至自身也难以理解。这表明社会共享的集体解释资源中针对某些群体的社会经验存在不足，这种不足导致集体对这一经验的理解缺失，从而剥夺了亲历者重要的自我理解。这种伤害的本质在于社会权威观念通过集体解释资源建构个人身份，使个体被建构为与真实身份不符的状态，从而阻碍其利益与发展。（[8]，p.168）此外，解释非正义还会引发一系列次要伤害，既包括与更系统的社会非正义相联系的实践层面的劣势，比如工作机遇、个人发展、身心健康等，还包括在认知层面阻碍主体获取新知识，甚至导致集体知识的损失。综上，解释非正义的发生包含了三个主要条件：（1）解释边缘化的存在；（2）说者某些重要的社会经验被掩盖，从而阻碍了理解和交流；（3）产生系统性伤害。

社会科学通常从两方面塑造社会的集体解释资源：第一、人们常利用社会科学知识来理解和解释自身处境，比如借助心理学理论认识并缓解自己的心理问题；第二、公共部门将社会科学视为公共政策制定的重要依据，这为社会科学将研究中隐含的结构性非正义投射到制度所处的规范领域提供了渠道。[28] 公共政策的制定本身也在创造社会解释资源，比如国际组织定义了共同的国际任务（如“发展”），创造并定义了新的行为者类别（如“难民”），并在世界各地转变了政治组织模式（如“市场”和“民主”）。[29] 在当代社会科学研究中，大数据以前所未有的规模构建了新的社会科学知识生产模式，改变了与之相关的社会政策和认知实践，进而塑造了社会的集体解释资源。基于这一前提，可以依次考察当代社会科学研究中源自大数据傲慢的解释非正义发生的三个条件：

第一，数字边缘群体遭受解释边缘化。随着在线数据量的爆炸式增长，当代社会科学研究越来越依赖机器学习来从海量数据中识别和组织信息模式。机器学习算法倾向于将统计上占主导地位的信息视为认知上相关的，而在处理与主流模式不一致的信息时存在结构性缺陷。如女性主义视角下的数据反思所揭示的：主导的大数据话语往往会产生以健全男性、异性恋规范的、中产阶级的及西方为中心的主体性和身份表现。[30] 大数据傲慢倾向于强化和复制在线数据中的主导模式，导致弱势群体观点被边缘化。阿帕杜莱（Arjun Appadura）指出，在当代英语国家的大数据媒体研究中存在一种将欧美的高级理论（high theory）与军事工业的算法高科技（high technology）相结合的特殊现象，从而导致认识论上的“东方化”（Orientalising），即暗示一种将东方视为他者、异域化和刻板化的过程。他将这种现象视为一种基于知识的帝国主义形式，通过普及西方的“基于机器的社会性”（machine-based sociality）的自动化聚合，来实现一种普遍性。[31] 这种大数据媒体研究不仅仅是在技术上的自动化和算法化，还是在认识论和文化层面将西方的视角和价值观强加于全球，导致非西方社会在数字世界中被边缘化。这种结构性缺陷就会导致在解释有关边缘化群体相关信息时面临概念资源的不足，从而使数字边缘群体在社会科学对大数据主导话语模式的强化与复制下遭受解释边缘化。

第二，数字边缘群体的某些重要的社会经验被掩盖。由于在线资源的腐蚀，大数据所介导的认知环境常常是“有毒的”，反映了主导群体的偏见和歧视性世界观，致使少数群体语言、规范和意义体系遭到系统性忽视。在这种情况下，数字化生活中的弱势群体与边缘群体往往难以被平等纳入社会科学的大数据研究中。社会科学理论实际上是社会不平等条件下特定阶级观点的代表，会局限于特定的“解释热点”，即那些“在社会生活中经适当诠释后对有权者有益的位置，或是在现存的错误解释

下维系有权者利益的位置”。（[8]，p.152）在这些“解释热点”上，由于具备科学权威与公信力的社会科学对集体社会意义的塑造，弱势群体的经验会被“积极地”掩盖。欧盟全球移民方法（EU's Global Approach to Migration）试图通过大数据分析追踪和预测移民人口流动性，以期更协调一致地解决移民问题。这种基于大数据的形式化严重依赖以往的偏见预设，将迁移现象抽象化为可预测的风险，将“移民”简化为可以被统计和控制的对象，而未能考虑他们复杂的背景、动机以及在迁移过程中的脆弱性和应对能力。[32] 如果社会科学研究者因大数据傲慢而忽视大数据研究中的偏见问题，那些与数字弱势群体相关的重要社会经验就会在社会科学理论的局限性之下被遮蔽。

第三，数字边缘群体遭受认知与社会层面的系统性伤害。解释非正义的主要伤害在于集体解释资源中的权威观念会剥夺弱势群体理解并表达自身经验的能力，从而影响其自我认同。社会科学理论作为集体解释资源的重要组成部分，具有身份建构力量。这种力量一方面来源于其客观性标签所赋予的公信力，另一方面则通过转化为公共政策将理论中的不公制度化，而我们总是通过管理控制我们生活的实践被创造。[28] 因此，在这种身份建构力量的作用下，遭遇解释劣势的数字边缘群体由于被剥夺理解自身重要经验的能力而被社会权威观念所塑造，进而影响自我认同。此外，社会科学对大数据偏见的忽视也将使那些被排除在大数据之外的群体面临更广泛的社会非正义。例如，国内生产总值的计算常忽略女性的无偿劳动（如家务与照护工作），导致发展政策系统性忽视女性贡献，加剧性别不平等，造成“看不见的女性”现象。[33] 欧盟将移民视为“风险”的大数据分析也不仅会加剧社会对移民身份的负面看法，还会导致政策更多关注如何控制移民，而非更好保护移民权利。[32] 因此，社会科学研究者的大数据傲慢带来了两个主要问题：一是它伤害了数据弱势群体的自我认同，对其够成系统性伤害；二是它在知识获取和解释方面造成了结构性缺陷，对社会科学自身造成伤害，进而可能导致基于这些研究的公共政策出现不公正，强化歧视性的社会结构。

综上，当代社会科学通过数据挖掘和统计分析来识别信息的模式和趋势，以实现预测和因果识别。然而，这种依赖于统计频率的研究方法导致了一种特定的“认知从众主义”，即研究倾向于只将在统计上占主导地位的信息视为认知相关，主要关注训练数据中频繁出现的模式，而忽视并排斥统计上较弱的信息。这种偏向使弱势群体的非主流信息和知识被边缘化，从而难以在公共知识系统中得到体现和认可。这不仅削弱了弱势群体的知识创造和传播能力，而且剥夺了他们参与社会认知实践的机会。作为社会解释资源的重要来源，社会科学研究结论在公共决策中发挥着关键作用。因此，如果社会科学由于大数据傲慢忽视未纳入大数据集的边缘化群体的观点或行为，这些群体就会在利用社会解释资源时处于劣势地位，同时在以社会科学研究结论为依据的公共决策中遭受更系统的社会非正义。

三、认知谦逊作为可能的应对方案

弗里克尝试提出一种认知德性（epistemic virtue）克服人们日常认知活动中身份偏见的影响。那么，社会科学研究者应该具备怎样的认知德性，以促进更加公正的社会科学研究？针对解释非正义的认知障碍，弗里克所提出的认知德性要求主体需要表现出一定程度的宽容或耐心，意识到说者此时表达困难甚至难以自我理解是客观层面上的解释鸿沟所致，而非说者主观上的认知失败。（[8]，p.169）这种反思批判意识确实可以减轻解释非正义对人们日常交流带来的负面影响。然而，解释非正义是一种结构性的非正义，弱势群体的解释边缘化地位植根于更广泛的社会非正义。实际上，弗里克诉诸于个体主义的认知德性方案遭到了来自结构主义的质疑，结构主义者并不否认个人认知努力的价值，但认为其并不足以应对社会结构问题。[34]

沃德罗普（Alistair Wardrope）在面对医疗

化（medicalization）现象所导致的解释非正义时，提出一种认知谦逊的解决方案。[35] 尽管认知谦逊仍难以完全解决导致解释非正义的结构问题，但其发展了弗里克的认知德性，既包括对反思批判意识的要求，又在此基础上强调了整合多元视角的开放性。医疗化现象是指将行为、心理现象和身体经验纳入医学定义、解释和治疗的过程。批评者认为这一过程过度强调生物学基础的解释，而忽视了生物医学模型之外的社会文化因素。为了应对生物医学模型的认知特权所导致的解释非正义，认知谦逊承认某人或某种视角的认知特权在一些研究领域的合法性，但要求主体必须警惕与这种认知特权不相称的可能性，并向外寻求多种资源弥补解释上的缺陷。与医疗化现象导致的解释非正义类似，当代社会科学研究中解释非正义的根源在于研究共同体赋予了数据技术不恰当的认知特权，因此认知谦逊可以削弱数据技术在社会科学中的过度主导地位，促进知识生产的多样性与公正性。这不仅是对个体研究者的要求，更是一种对社科共同体的制度性呼吁，要求社科共同体承担相应的集体认知责任。

根据认知谦逊的要求，社会科学共同体在研究中首先需要培养更加敏锐的反思批判意识，以识别和修正大数据中的潜在偏见，并公开讨论其局限性。朗基诺（Helen Longino）在《作为社会知识的科学》（*Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*）中提出，科学共同体可以通过批判性的对话和开放的辩论来检查和评估科学假设和科研数据，揭示隐藏的偏见，并鼓励科学家对研究结果进行多样化的解释。她认为，在一个既定的科学共同体内存在越多不同的观点与意见，这个共同体对自然过程的描述与解释就可能更多地是根据事实，其研究活动、研究成果就越客观；对研究成果批判的程度和范围决定客观性的程度，批判的程度越强，批判的范围越宽，客观性就越强。[36] 面对大数据在当代社会科学研究中日渐形成的“认知权威”，反思批判意识要求社会科学共同体克服自身“认知惰性”，在分析数据时能主动意识到数据驱动的知识生产实践植根于特定情境，研究中所利用的数据是不完整、不自明的。认识并反思数据技术的偶然性及其认识论特征，将使社会科学能够产生更健全和有意义的知识。[30]

认知谦逊还要求社会科学共同体寻求大数据之外的认知资源以减轻偏见问题。具体而言，在数据收集过程中，可以通过三角测量方法整合多种视角与信息减少大数据中可能存在的偏见。三角测量方法是指通过多种途径、方法或数据源来对同一研究对象或现象进行测量与观察，以提高研究结果的可靠性和有效性。在数据分析过程中，研究者应更主动地将地方知识和更具情境化的解释融入进更具包容性的数据分析方式中。鲁尔斯（Koen Leurs）在其有关数字身份的研究中增加了与伦敦年轻人面对面访谈的方式，与其共同制作了脸书友谊网络的可视化图表。通过这些方法，参与研究的年轻人不仅决定了他们想要分享哪些数据，还作为共同研究者参与其中，从而更准确反映其生活经验和观点。[30] 弗雷（William Frey）等也通过“跨界合作”让曾经参与帮派的年轻人作为“专家”参与到芝加哥帮派青年的推特数据解释中，这种方法有助于避免研究者断章取义的预测，从而减少了对芝加哥青年的误判。[37] 以上方法实际上都是通过扩展社会科学共同体的“认知多样性”帮助识别和纠正大数据中隐含的偏见。多样性是批判性观点的来源，不仅有助于科学家识别和纠正错误的背景假设，还能迫使科学家为其观点提供更强的论据，避免陷入教条主义。

随着社会科学研究越来越依赖于数据驱动的人工智能系统，这些系统基于大量历史数据识别模式，从而影响对未知的预测和探索。这种对“已给定”的无批判接受可能会加剧知识和权力不平等，强化科学研究中的主导观念，限制科学创新和多样性。此外，其“黑箱”特性削弱了科学探索的根本特征——持续的质疑和开放性探讨，威胁到科学的民主本质与人类创新能力。认知谦逊要求研究者具备反思批判意识，更强调一种开放的研究态度。在利用大数据、人工智能等新兴技术时，研究者应保持

谨慎，避免盲目追求技术而忽视研究的本质。同时，研究者还需要意识到数据局限性，并寻求其他数据源和研究方法来补充和验证研究结论。然而，社会科学研究者在寻求互补资源和整合多元视角时，可能会遇到资源获取困难、不同观点难以协调等实际问题。尽管如此，认知谦逊无疑是当代社会科学研究伦理的内在要求，其不仅要求社会科学共同体推进知识的发现与传播，更要求其对知识的后果承担集体认知责任，容纳更广泛的批判性视角。

结　语

本文考察了当代社会科学研究中因大数据傲慢引发的解释非正义问题，并提出认知谦逊作为可能的应对方案。随着社会科学研究进入到数据密集型知识发现的新阶段，海量数据及相关的分析技术正在超越从解释到预测的传统方式，更有效地揭示变量之间的相关性，从而做出更精准的预测，帮助人们深化对重大社会问题的认识。这使得社会科学研究在社会生活和公共治理中发挥日益重要的作用，其研究成果常常会对政策制定和社会发展产生深远影响。然而，大数据傲慢却削弱了社科研究者对研究中伦理问题的敏感性，对大数据偏见的忽视可能导致弗里克所提出的解释非正义，使弱势群体的声音被系统性忽视，加剧了社会不平等。因此，社会科学共同体应当在研究过程中纳入更多的伦理考虑，充分审视大数据的广泛运用是否出于自身的某种“认知惰性”，谨慎对待大数据可能形成的“认知权威”，积极寻求数据和技术的“认知多样性”。大数据并不等于客观性，相反，对大数据的过度依赖可能会破坏社会科学知识的客观性和“充分代表性”。社会科学共同体的公开性和包容性会影响理论的客观性和合理性，一个越具包容性和公开性的共同体，其探究活动就更有可能产出客观的知识。

［参 考 文 献］

[1]Hey, T., Tansley, S., Tolle, K. M. 'Jim Gray on eScience: A Transformed Scientific Method'[A], Hey, T., Tansley, S., Tolle, K. M., et al. (Eds.) *The Fourth Paradigm: Data Intensive Scientific Discovery*[C], Washington: Microsoft Research, 2009, xvii–xxxi.

[2]Bertoni, E., Fontana., M., Gabrielli, L., et al. (Eds.) 'Mapping the Demand Side of Computational Social Science for Policy'[R]. Luxembourg: Publications Office of the European Union, 2022.

[3]Hollingshead, W., Quan-Haase, A., Chen, W. H., et al. 'Ethics and Privacy in Computational Social Science: A Call for Pedagogy'[A], Engle, U., Quan-Haase, A., Liu, S. X., et al. (Eds.) *Handbook of Computational Social Science: Volume 1*[C], London: Routledge, 2021, 171–185.

[4]Bertoni, E., Fontana, M., Gabrielli, L., et al. (Eds.) *Handbook of Computational Social Science for Policy*[M]. New York: Springer, 2023.

[5]凌昀、李伦. 计算社会科学研究：范式转换与伦理问题[J]. 江汉论坛, 2020,（9）: 26–31.

[6]Lazer, D. M., Pentland, A., Watts, D. J., et al. 'Computational Social Science: Obstacles and Opportunities'[J]. *Science*, 2020, 369(6507): 1060–1062.

[7]Stommel, W., Rijk, L. 'Ethical Approval: None Sought. How Discourse Analysts Report Ethical Issues Around Publicly Available Online Data'[J]. *Research Ethics*, 2021, 17(3): 275–297.

[8]Fricker, M. *Epistemic Injustice: Power and the Ethics of Knowing*[M]. Oxford: Oxford University Press, 2007.

[9]Pongiglione, F. 'Epistemic Hubris'[J]. *Social Epistemology*, 2025, 39(1): 91–105.

[10]Boyd, D., Crawford, K. 'Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon'[J]. *Information, Communication & Society*, 2012, 15(5): 662–679.

[11]Lazer, D. M., Kennedy, R., King, G., et al. 'The Parable of Google Flu: Traps in Big Data Analysis'[J]. *Science*, 2014, 343(6176): 1203–1205.

[12]Messeri, L., Crockett, M. J. 'Artificial Intelligence and Illusions of Understanding in Scientific Research'[J]. *Nature*, 2024, 627(8002): 49–58.

[13]Uprichard, E., 'Most Big Data is Social Data: The Analytics Need Serious Interrogation'[EB/OL]. http://blogs.lse.ac.uk/impactofsocialsciences/2015/02/12/philosophy-of-datascience-emma-uprichard/. 2015–02–12.

[14]罗伯特·毕夏普. 社会科学哲学：导论[M]. 王亚男 译, 北京：科学出版社, 2017, 277–285.

[15] 米加宁、章昌平、李大宇 等. 第四研究范式: 大数据驱动的社会科学研究转型 [J]. 学海, 2018,（2）: 11–27.
[16] 张庆熊. 大数据时代社会科学方法论探讨 [J]. 社会科学, 2018,（9）: 69–77.
[17] 维克托·迈尔–舍恩伯格、肯尼思·库克耶. 大数据时代 [M]. 盛杨燕、周涛 译, 杭州: 浙江人民出版社, 2013.
[18] 倪万、唐锡光. 大数据应用于社会科学研究的价值与悖论 [J]. 东南学术, 2017,（4）: 68–78; 247.
[19] Kitchin, R. 'Big Data, New Epistemologies and Paradigm Shifts'[J]. *Big Data & Society*, 2014, 1(1): 1–12.
[20] Couldry, N., Mejias, U. A. 'Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject'[J]. *Television & New Media*, 2019, 20(4): 336–349.
[21] Gitelman, L. (Ed.) "*Raw Data" is an Oxymoron*[C], Cambridge: The MIT Press, 2013.
[22] 方兴东、钟祥铭. 中国数字治理发展报告 [M]. 北京: 社会科学文献出版社, 2023.
[23] Robinson, L., Ragnedda, M., Schulz, J. 'Digital Inequalities: Contextualizing Problems and Solutions'[J]. *Journal of Information, Communication and Ethics in Society*, 2020, 18(3): 323–327.
[24] Hargittai, E. 'Potential Biases in Big Data: Omitted Voices on Social Media'[J]. *Social Science Computer Review*, 2020, 38(1): 10–24.
[25] 马克·里斯乔德. 当代社会科学哲学: 导论 [M]. 殷杰、郭亚茹、申晓旭 译, 北京: 科学出版社, 2018, 27–29.
[26] 瑞文·康奈尔. 南方理论: 社会科学知识的全球动态 [M]. 詹俊峰 译, 南京: 江苏人民出版社, 2024.
[27] Sharifzadeh, R. 'Digital Methods: An STS Challenge to Methodological Digitization in Social Science Research'[EB/OL]. *Social Epistemology*, 2024, 1–12. doi: 10.1080/02691728.2024.2400565.
[28] McCollum, J. 'Hermeneutical Injustice and the Social Sciences: Development Policy and Positional Objectivity'[J]. *Social Epistemology*, 2012, 26(2): 189–200.
[29] Barnett, M., Finnemore, M. 'The Politics, Power, and Pathologies of International Organizations'[J]. *International Organization*, 1999, 53(4): 699–732.
[30] Leurs, K. 'Feminist Data Studies: Using Digital Methods for Ethical, Reflexive and Situated Socio-cultural Research'[J]. *Feminist Review*, 2017, 115(1): 130–154.
[31] Appadurai, A. 'The Academic Digital Divide and Uneven Global Development'[J]. *CARGC Papers*, 2016, (4): 1–10.
[32] Taylor, L., Meissner, F., 'A Crisis of Opportunity: Market-making, Big Data, and the Consolidation of Migration as Risk'[J]. *Antipode*, 2020, 52(1): 270–290.
[33] Anderson, E., 'Pragmatism, Science, and Moral Inquiry'[A], Fox, R., Westbrook, R. (Eds.) *In Face of the Facts: Moral Inquiry in American Scholarship*[C], Cambridge: Cambridge University Press, 1998, 10–39.
[34] 谭笑. 个人德性的解决方案是否切实有效?——从情绪与认知角度来看 [J]. 自然辩证法研究, 2023, 39（2）: 60–66.
[35] Wardrope, A. 'Medicalization and Epistemic Injustice'[J]. *Medicine, Health Care and Philosophy*, 2015, 18: 341–352.
[36] Longino, H. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*[M]. Princeton: Princeton University Press, 1990.
[37] Frey, W. R., Patton, D. U., Gaskell, M. B., et al. 'Artificial Intelligence and Inclusion: Formerly Gang-Involved Youth as Domain Experts for Analyzing Unstructured Twitter Data'[J]. *Social Science Computer Review*, 2020, 38(1): 42–56.

[责任编辑 李斌]