

# 谁能从事计算社会科学研究

## ——数据增值过程与学术分层的耦合机制

牛朝佳<sup>1</sup> 郭台辉<sup>2</sup>

**【内容摘要】** 近年来，数据驱动的计算社会科学发展迅速，而各科研主体对数据资源的占有与处理能力存在差异，由此产生知识生产的不平等问题，即谁能从事计算社会科学研究？在拉图尔“计算中心”概念基础上，可以构建数据形态的分析框架，即将数据视为要素、资本与符号，对应计算社会科学知识生产过程的进入、角逐与垄断阶段。在进入阶段，数据作为生产要素成为科研门槛，基础设施与数字技能构成进入壁垒；在角逐阶段，数据要素转化为数据资本，不同主体围绕数据展开合作与应用，导致资源与权力逐渐向中心集聚；在垄断阶段，数据资本经由传播与奖励体系符号化，转化为话语权与评价权，进一步固化知识生产中的等级结构。即便如此，未来的计算社会科学仍有可能迈向共建与共享的多元主体格局。

**【关键词】** 计算社会科学 数据要素 数据资本 数据符号 计算中心

**【作者】** 1 牛朝佳，云南大学政府管理学院博士研究生；  
2 郭台辉，云南大学民族学与社会学学院教授。（昆明 650091）

### 背景与问题

作为一个新兴研究领域，计算社会科学以前所未有的广度、深度和规模来收集与分析数据，揭示个体和群体行为的模式。<sup>①</sup>以拉泽尔等人 2009 年在 *Science* 发表的文章为标志，计算社会科学迅速发展，多种专业期刊创立，研究机构和学术团体纷纷成立，论文发表数量逐年增加（见图 1）。全球不少高校相继开设“计算社会科学”专业，中国一些重点高校也成立“计算社会科学实验室”。<sup>②</sup>

计算社会科学与传统定量社会科学的最明显不同在于数据来源与体量。在过去一百多年的小



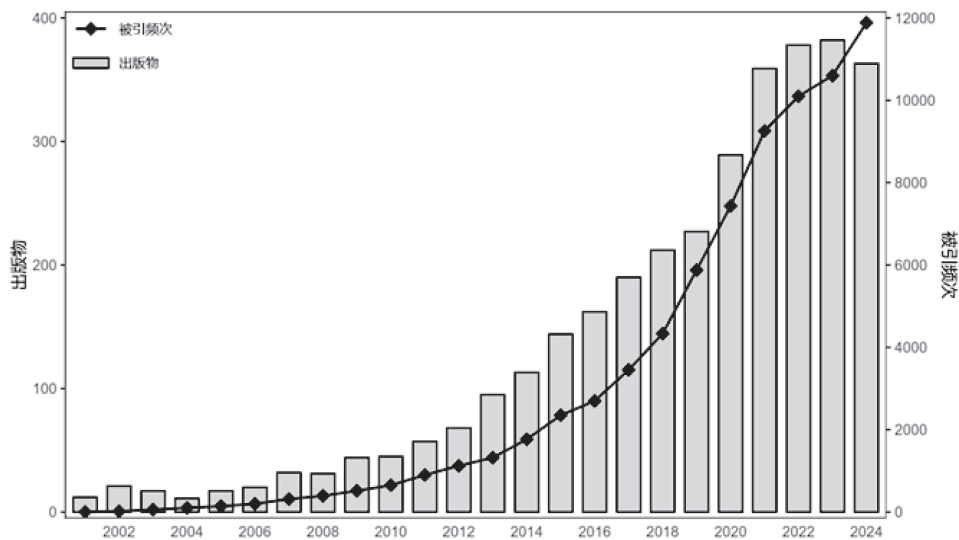


图1 与计算社会科学相关的出版物与被引频次分布 (2001—2024)

数据来源：Web of Science 核心合集。主题 = Computational Social Science，出版年 = 2001—2024。共检索出 3289 条文献，总被引频次 75473。检索时间为 2025 年 3 月。

数据时代，定量社会科学是通过“假设—验证—结论”的基本程序，借助社会调查和统计资料得来“闭源数据”，据此提供对社会现象的模式解释。但在计算社会科学时代，“开源数据”成为学者们的分析对象，他们弱化模型的设定与假设，收集互联网自然保留的海量数据，发现其中的有效信息，客观上可以提供更多的模型选择空间和启示。<sup>③</sup>然而，数据驱动的计算社会科学研究也面临一些挑战。比如数据访问受限，数据共享机制缺乏，跨学科合作体制不健全，人才培养机制尚未建立，基础性计算设施不完善等，而数据问题被视为限制计算社会科学持续发展的核心瓶颈。<sup>④</sup>拉泽尔等人进一步指出，计算社会科学可能成为私人公司和政府机构的专属领域，导致一部分学者享有更多特权，他们获得私人数据的控制权，并基于这些数据从事科学研究，但这些研究成果无法批判或复制。<sup>⑤</sup>这些问题导致公众质疑计算社会科学的科学性，还引发严重的马太效应。换句话说，拥有数据优势的机构和团队将不断积累更多成果，获取更多的研究资金和社会政治关注，反过来，这增加他们对数据要素的支配权，使他们进一步垄断知识生产，而占有数据要素较少的研究者则愈加边缘化。

这种数据驱动的学术不平等现象，在知识生产的每一个环节都因数据形态的演化而不断扩大。首先，只有具备完善的算力集群和爬取、清洗、存储等硬件—软件条件，才能批量搜集社交媒体流、行政记录等海量原始数据，形成可供科研使用的数据要素。这道门槛决定了谁能进入计算社会科学领域。随后，数据要素在跨学科、跨机构乃至企业（政府）—学术联盟的协作中被资本化，成为可交换、可增值的数据资本。强势团队借此在项目申请、经费分配和成果转化方面进行竞逐，并扩大与其他研究者的差距。当这些团队的科研成果借由高被引论文、学术会议与媒体平台得以扩散，并获得基金奖项或学术头衔时，数据进一步上升为象征性权威的数据符号，这标志着一些学术团队已经获得议题设定的话语权与资源分配的领导力。

那么，数据是如何沿着“要素—资本—符号”的增值过程累积优势，并通过“基础设施—合作应用—传播奖励”体系不断扩大学术差距的？为此，本文搭建数据形态的分析框架，从“进入—

竞逐一垄断”三个阶段出发，考察数据形态的演化如何在计算社会科学研究中形成学术分层，进一步讨论计算社会科学得以持续发展的可能方向，并主张建设一个开放共享的新兴领域。

## 数据形态的分析框架

进入数字时代，人类大部分行为都留下数字化的痕迹。这些痕迹被收集并归档，形成具有庞大信息量的开源数据库，无论是商业还是非商业组织都越来越关注并挖掘其无穷的资源与价值。<sup>⑥</sup>因此，数据逐渐成为一种新型的战略资源，被称为“新石油”。<sup>⑦</sup>谁能够在数据资源上占据优势，谁就更容易在数字时代的竞争中取得领先地位。<sup>⑧</sup>计算社会科学是一门数据驱动的、以数据密集化为特征的新兴交叉学科，其研究起点是占有数据资源。研究者以真实世界为对象，依赖先进的技术手段获取数据，并运用工具进行统计、计算和内容分析。<sup>⑨</sup>但是，这类数据相较于传统社会科学数据更嘈杂、更无结构且更缺乏“设计”成分，使得处理该类数据需要借助计算机科学领域的方法和技术。<sup>⑩</sup>

计算社会科学强调获取在地化信息、构建关联的全局性，<sup>⑪</sup>从而符合拉图尔提出的“计算中心”（centres of calculation）概念。<sup>⑫</sup>计算社会科学的核心是数据挖掘，其本质是知识发现的过程。<sup>⑬</sup>然而，并非所有参与主体都能够有效利用这些数据并将其转化为知识。在拉图尔看来，知识生产是一个累积循环的整体，即把一切采集物累加集中于某一节点并不断循环此过程，新要素便会不断聚集其中，该节点也就逐渐成为知识生产的中心。这表明，在计算社会科学领域中，那些拥有数据收集、分析和解释能力的研究主体，将成为知识生产的中心，并在竞争中占据结构性优势。

然而，“计算中心”并非自然形成，而是在持续不断地收集、存储和分析数据的累加循环过程中得以塑造和巩固。因此，有必要聚焦数据的三种形态，考察计算社会科学领域的知识生产过程：数据要素、数据资本和数据符号。这三种数据形态并非彼此孤立，而是随着数据的价值增值连续演进。此外，围绕数据资源的占有与控制，计算社会科学领域的参与主体表现为进入、竞逐与垄断三种状态。该过程的每一个阶段都导致参与群体的内部分化，最终结果是，少数处于“计算中心”的研究主体获得话语权与支配权，而大多数研究主体因缺乏关键的数据资源和相应技术能力而边缘化，逐渐被排除于数据驱动的社会科学知识生产过程之外。

首先，数据作为要素的进入阶段。未经数字技术开发的数据只是静态的数据资源，既不能充分释放数据的自然价值，也不能实现数据的社会价值。<sup>⑭</sup>参与主体有目的有组织地收集、存储和处理这些广泛存在于互联网中的原始数据，才能将其转换为一种具有生产力的要素，即数据要素。然而，这一数据转化过程要求大量硬件与软件的投入，以及具备相关数字技能的专业科研人员参与，这成为潜在参与主体的门槛。具体而言，算力、算法和掌握数字技能的人才共同构成计算社会科学研究的设施，形成准入的门槛与壁垒，决定了哪些参与主体能够进入并实际开展研究。

其次，数据作为资本的竞逐阶段。当原始数据通过基础设施转化为数据要素后，其内在价值逐渐显现，形成数据资本。数据资本是以现代信息网络和各类型数据库为重要载体，基于信息和通信技术而充分数字化、生产要素化的信息和数据。<sup>⑮</sup>这种特性导致数据价值实现过程不是简单的占有数据要素，而是依赖于不同参与主体的合作与应用，实现数据的整合和价值增值。

数据生成过程通常涉及多个主体，包括产品供需双方、平台企业及政府机构等，这使得数据自生成之时便分布于不同主体。<sup>⑩</sup>因此，只有参与主体具备完善的基础设施、能有效整合不同来源的数据并持续地与其他主体合作，才能实现数据资本的价值最大化。然而，这一过程所依赖的资源与能力门槛更高，这使得不同主体在数据分析、解释和应用等方面呈现显著的不对称。那些缺乏足够数据要素、技术应用能力与合作网络的参与主体便逐渐丧失竞争力，并逐渐被排除于计算社会科学研究之外。由此可见，数据资本化的过程加剧了计算社会科学领域内部的不平等与分化。

再次，数据作为符号的垄断阶段。布尔迪厄认为，任何形式的资本只要在社会情境中被“认可”，就可以转化或表现为象征资本。<sup>⑪</sup>伴随着数据资本的积累与增值，数据资源不仅成为合作与竞争的核心对象，更在制度化的传播与奖励体系中获得象征性的权威地位。在计算社会科学领域，这一过程主要通过著作、论文、基金项目、科学奖项和业界应用等传播与奖励体系实现。那些能够持续获得主流传播渠道认可和高等级学术奖励的参与主体，其数据分析结果和知识产出更容易被视为标准、权威乃至范式，从而赋予其显著的象征资本。这种象征资本不仅反映为学术声望的累积，还表现为话语权、标准制定权乃至研究资源配置权的再生产。与此同时，这又反过来增强其获取数据要素的能力，扩大其合作伙伴的范围，进一步提高其声誉。而那些无法获得主流传播和奖励体系认定的研究团队或机构难以被计算社会科学主流群体承认，便逐渐被边缘化。正是通过传播与奖励，数据资本得以符号化，并最终固化为计算社会科学领域内的等级结构与权力垄断。在这个意义上，谁拥有足够多的数据资源，谁就拥有权力；谁拥有权力，谁就能拥有更多的数据。<sup>⑫</sup>

综上，数据形态的分析框架逐渐清晰（见图2）。首先，数据作为要素，在基础设施的支持下实现原始数据的收集与转化，为知识生产提供基本支撑。其次，数据在合作与应用的过程中通过多主体整合和技术赋能，进一步演化为可积累、可增值的资本，成为竞争与协作的核心资源。最后，数据资本经由制度化的传播与奖励机制符号化，成为衡量学术地位、分配话语权与构筑知识分层的重要依据。这一转化路径不仅加剧了计算社会科学领域内部的等级化趋势，也促使我们思考一个问题：在数字时代的知识生产过程中，谁更能获得主导权与话语权，谁更可能被逐步边缘化？

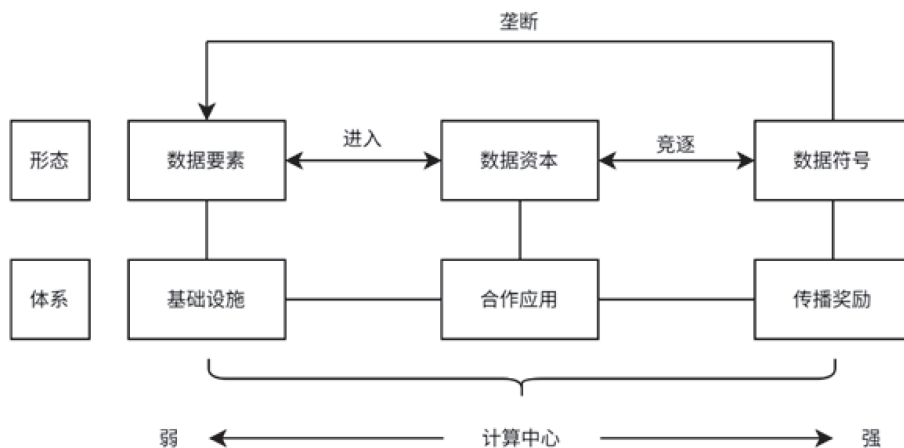


图2 数据形态的分析框架

## 进入：数据要素与基础设施

在数据驱动的计算社会科学中，大数据（比如社交媒体数据、政务公共数据、影视书籍资料等）是开展实证分析和模型计算的基础，其质量直接决定了研究者能够提出的问题类型和采用的方法深度。因此，从事计算社会科学研究起点是对原始数据的收集。数据是指利用一定信息技术而形成信息的数字化记录，<sup>①</sup>它不仅可以帮助企业降低成本、提高效率，还可以推动学术研究和知识积累。<sup>②</sup>数据可分为结构化数据与半结构化数据，前者可直接分析应用，而后者需经过一定步骤将其转换为结构化数据才可使用。但是，数字社会中的大多数数据都以非结构化的形式存在。因此，为了获取这些原始数据，参与者要投入资本与劳动，通过设置收集机制，捕获数据字节，并转换为结构化数据，由此，将生产中的附加价值添加到原始数据的价值中，使之成为一种具有生产力的数据要素。<sup>③</sup>

与劳动、资本、土地、知识技术和管理一样，当数据与生产资料分离时，它们只是可能的生产要素。要成为现实的生产要素，数据必须与劳动、技术等相结合。<sup>④</sup>数据价值链反映了数据增值所经历的环节，一些学者对数据的价值链进行不同的阶段划分，比如“采集或授权”“储存和聚合”“分析和销售”和“利用”四阶段，<sup>⑤</sup>或者“采集”“存储”“加工”“销售”和“利用”五阶段，<sup>⑥</sup>或者“数据收集”“数据存储”和“数据分析”三阶段，<sup>⑦</sup>或者“数据采集”“数据清洗与存储”和“数据加工”三阶段。<sup>⑧</sup>借鉴上述研究，本文将原始数据形成数据要素的价值链更简单地划分为数据收集、数据清洗与存储两个阶段。经此不断积累原始数据的价值，形成数据要素。

在数据收集阶段，研究者有意识、有目的地将各个社会场景中的复杂信息以字符串的形式记录下来，得到原始数据。原始数据虽然广泛存在，但无法随意捕捉。一方面，尽管数据的收集与使用不影响其他市场主体对该数据的利用，<sup>⑨</sup>但数据的获取技术与资源仍存在壁垒，这是因为收集可利用的数据需要数目可观的投资，<sup>⑩</sup>包括对劳动技能的投资与建立收集数据所需的软硬件设施的投资。这些前期投资并非所有科研机构可以承担，导致一些机构无法从事计算社会科学研究。另一方面，由于数据隐私与伦理问题，许多数据被平台企业与政府部门垄断，无法通过公开渠道获取。只有那些声誉较高的科研机构可以通过签署合作协议的方式获得数据使用权，而那些无法与企业与政府建立联系的机构则无法获得数据，面临进入壁垒。

在数据清洗与存储阶段，研究者对原始数据进行清洗，去除无效和冗余信息，得到质量较高的、可用的结构化数据，并将其存储于云端或本地服务器，方便后续的检索和应用。这个阶段仍然需要投入劳动和资本要素，比如存储设备的投资、软硬件的投资等。

由上述价值链可见，数据转为数据要素源于每个环节所投入的成本，以及对数据加工技术的应用。围绕上述生产活动，参与主体构建的数据基础设施可概括为相应的资本投资、物理硬件和人力资源。只有拥有这些基础设施，研究者才能有效获取大规模数据。因此，数据基础设施在很大程度上决定了谁可以优先进入计算社会科学研究领域并对此进行垄断。一旦参与主体将尽可能多的数据要素集中起来，并大致规划其用途、目的、预期产出等，累加循环的马太效应便会发生。通过累加性的循环，一些参与主体逐步掌握大规模的数据要素，成为事实上的“计算中心”。

以谷歌学术为例，它收集并存储了海量的学术文献数据。平台的索引范围和算法排序虽对学

术研究具有巨大影响，却是不透明的，且不受学术界控制。这使平台能够在很大程度上决定用户获取知识的方式与深度。这一现象表明，数据基础设施不仅是一种技术装置，更是一种权力的载体，它可以决定哪些主体能进入知识生产的中心，哪些主体则被排斥在外。这种作为权力载体的数据基础设施，其核心特征体现为拉图尔所描述的“计算中心”。这些“计算中心”已然不是被动的数据存储库，而是主动收集、处理和转化信息的关键节点。它们通过标准化的计算流程，将分散且异质的原始数据转化为可比较、可利用的数据要素，为后续的进一步分析、应用和决策奠定了关键基础。

因此，对这些“计算中心”的控制，意味着掌握了信息流向、知识定义和议程设置的远程支配权。在计算社会科学领域，无论是商业平台搭建的庞大数据系统，还是科研机构创建的数据库，都在不同程度上扮演着“计算中心”的角色。通过架构设计、算法和访问规则，这些“计算中心”不仅筛选信息，更塑造研究者观察社会的方式，从而巩固特定主体的优势地位，并可能边缘化那些缺乏入场资格的参与者。因此，“计算中心”不仅代表信息处理能力的集中，更体现一种新的权力分配模式，将资源分配、规范制定与身份建构交织在一起，构成进入该领域的门槛。

显然，在数据要素的进入阶段，数据基础设施所形成的权力边界，不可避免地导致不同主体之间的结构性分化。资源相对匮乏的科研机构往往在数据获取和算力支持方面面临困境，它们或难以承担构建与维护先进基础设施的高昂成本，或在与掌握核心数据的平台企业及政府部门协商时处于弱势地位，导致研究活动受限。相比之下，已有显著声誉的科研机构凭借丰富的数据要素积累与强大的技术实力，能够构建私有且高度集成的“计算中心”。这些机构不仅能高效挖掘数据价值以实现科研目标，还能逐渐在知识生产领域获得话语权，从而形成事实上的“数据霸权”。因此，进入这一并非中性的技术准备过程，构成了学术分层的第一道门槛。也正是在此阶段，关于谁能从事计算社会科学的核心问题，其答案似乎已经初步显现：谁掌握了完善的数据基础设施，谁便初步拥有了研究的主导权以及对潜在参与者的支配权。

## 竞逐：数据资本与合作应用

数据收集为算法提供基础信息，而算法又反过来指导数据的进一步收集，二者相互促进。算法规模越大，即时收集和整理数据的能力也越强，<sup>⑳</sup>“计算中心”的作用随之增强。然而，拥有数据要素仅代表具备了将其转化为生产力的潜在可能，参与主体自身的数据分析能力及相关基础设施则是数据要素发挥作用、推动知识生产的必要前提。<sup>㉑</sup>简单堆积的数据要素并不能自动转变为数据资本。只有投入大量劳动、资金和技术，经过系统加工并可以为新生产过程提供支撑的数据要素，才能成为数据资本。<sup>㉒</sup>因此，数据资本不仅体现为价值链各环节的投入成本，也体现为对数据加工技术的应用程度。<sup>㉓</sup>由于数据资本是对数据要素的不断积累、加工得到的结果，该过程可以划分为数据分析、数据共享与应用两个阶段。

在数据分析阶段，“计算中心”的数据科学家或科研人员借助相关硬件设施，运用数据科学、数据挖掘和机器学习等手段，对已结构化的数据要素进行深入加工，从而提取有价值的知识、建立知识图谱，促进数据要素进一步增值。<sup>㉔</sup>一旦数据收集和存储的固定成本得到满足，数字服务的有用效果几乎可以无成本地利用。<sup>㉕</sup>但这种观点隐含了“一旦获得数据要素，其中的知识便可立即无成本地显现”的假设，忽略了知识生产的复杂性和成本。<sup>㉖</sup>在现实情景中，从数据中提取

知识需要进行深入的数据挖掘、处理分析和解释推理。<sup>⑥</sup>这必然要求参与主体拥有强大的数据分析能力，包括算力与算法的支持，以及掌握数字技术的科研人员。这一阶段的数据价值增值主要源于数据分析技术的融入，数据分析技术的进步也会直接提高数据要素的价值。<sup>⑦</sup>

在数据共享与应用阶段，处于“计算中心”的研究者基于数据分析，发表成果与制定决策，并进一步将其应用于具体的现实场景。反过来，这种实践刺激“计算中心”的各类主体扩大数据共享范围，加强合作分析，并将数据资本更有效地应用于更广泛的领域，发挥更大价值。在合作网络中，数据资本的生成依赖于多方主体的协同投入与共享机制。例如，平台企业可以向合作伙伴开放内部的私有数据接口，帮助研究人员获取新知识；反过来，学术研究所产出的成果可以提升平台对数据的利用效率，实现数据要素的再生产。这样的合作与应用循环，原始数据不断形成新的数据要素，并进一步衍生出新知识，完成数据资本的增值和再生产。可见，数据资本的形成并非单一主体独立完成，而依赖于合作与应用时的数据流动与积累，其价值随之被放大。

进而言之，上述两个阶段导致部分潜在的参与主体被排除在计算社会科学研究之外。一方面，具备数据分析优势的主体可以高效处理不同规模的数据分析任务，快速挖掘数据要素中的有效信息，并结合既有理论进行诠释，从而转化为可利用的知识。<sup>⑧</sup>另一方面，出于强化数据挖掘能力和掌控战略资源的需求，各类主体往往对同类型的数据源表现出浓厚兴趣，尽可能捕获各类数据，这使得平台企业拥有更丰富的数据储备，并在大数据技术应用方面占据优势地位。<sup>⑨</sup>科研机构往往依赖于平台企业提供的数据资源，因而处于相对弱势的位置，而平台企业倾向于向声誉更显著的科研机构开放数据，资源较弱的科研机构则受到限制。此外，还有一部分高质量的数据掌握在政府部门手中，这些部门同样选择性地开放数据，从而进一步将部分参与主体排斥在计算社会科学研究之外。

从上述价值链可以看出，数据要素转化为数据资本的过程，实际上就是在“计算中心”内部建造中心，<sup>⑩</sup>合作网络呈现明显的集中化趋势，形成核心—边缘的权力关系结构。掌握丰富数据资源和计算能力的主体成为网络的中心节点，汇聚多源数据，推动知识生产和技术创新。这些中心节点通过主导数据接口标准、格式规范与共享协议，进而控制合作网络的运作模式。在合作与竞争的交互过程中，数据资源逐渐集中于少数中心节点，使其进一步强化为更具影响力的“计算中心”。数据资本优势明显的主体则利用已有优势巩固并扩展核心地位，新进入的主体则难以改变这种权力结构。这种不均衡的格局，导致数据资本的再分配通常会向既有权力中心倾斜。因此，所谓竞逐，本质上是各参与主体围绕能够将数据流转化为战略资产的基础设施及其塑造过程所展开的数据资本所有权和使用权的争夺。

### 垄断：数据符号与传播奖励

布尔迪厄认为，任何形式的资本只要被他人认知并承认，便具备符号层面的价值与威望，从而产生无形的力量。<sup>⑪</sup>当数据资本在具体的社会场域中被认知和承认，并进一步获得象征意义时，便可转化为符号资本，即数据符号。在科学场域中，这种符号资本的形成不仅意味着经济与技术优势的象征化，也意味着对话语权与评价标准制定权的掌控。数据资本的符号化并非自发实现，而是依赖于一整套制度化的机制，包括学术发表、同行评议、媒体传播以及荣誉奖励等。只有被主流学术共同体认可并赋予可信度的研究成果，才能转化为具有象征意义的数据符号，进而提升



研究者在科学场域的权威地位。

首先，学术出版与同行评议等传播机制扮演了“过滤器”的角色。只有经过严格审查、验证并被权威期刊接收的数据成果，才可能获得学术共同体的信任。这一制度性流程保障了数据成果的可靠性和有效性，为数据资本的符号化提供了必要的信誉基础。此外，学术共同体内部的信任网络也进一步强化数据的可信性。正如拉图尔等人指出的，在科学实践中，研究者的个人可信度与其研究结果的可信度密切相连，同行最关心的是研究者本人是否可靠，以及其研究结论是否令人信服。<sup>④</sup>研究者的学术履历与成果记录（如论文发表情况、职称头衔）成为这种可信性的主要载体，为数据成果提供制度性的信用。通过上述传播机制与信誉网络，数据成果的可信性最终得以被主流学术共同体认可和强化。

数据成果在获得可信性之后，还需要通过科学奖励体系转化为具有等级区分的符号资本。科学领域内的各种奖励（如荣誉称号、奖项、基金资助等）不仅是对科研贡献的肯定，更是一种象征性的认证，使数据成果获得不同层级的声望价值。默顿早在20世纪70年代便指出科学场域存在“马太效应”，即名声与荣誉具有累积增益的现象。<sup>⑤</sup>一旦研究者获得某项重要奖励并赢得一定知名度，就更容易获得更多荣誉和资源，从而形成优势的持续累积；而缺乏声誉者则愈发难以获得承认，陷入劣势的持续累积。因此，科学奖励制度制造出等级化的符号格局：顶尖学者由于频繁获奖，在学术场域中占据更高的象征地位，而新进入者或边缘学者则很难突破这一壁垒。这种由奖励机制推动的象征资本分化，使得数据成果及其相关研究者被赋予不同等级的声望标签，进一步强化了学术界的等级秩序。<sup>⑥</sup>在这一过程中，马太效应助长了“强者恒强”的局面：已有象征资本者持续获得超出其实质贡献的关注和褒奖，而欠缺名望者即便拥有卓越成果，也可能被忽视。因此，学术共同体内的奖励结构将数据成果转化为社会认可的符号标志，为日后象征资本的形成奠定基础。

数据资本转化为数据符号后，即获得象征性的学术声望，对科学场域的影响不止于个人荣誉层面，更在于对场域内部规则和权力结构的再生产。布尔迪厄指出，学术场域内部存在激烈的符号权力斗争，参与者竞争的目标在于垄断对场域内“合法性”的定义权。<sup>⑦</sup>拥有数据符号的科学家往往也是场域规则的制定者，能够掌控学术评价标准、话语权和研究议程，例如哪些研究问题被视为重要、哪些方法论被视为规范。这些规则和标准通常体现出场域中支配群体的价值取向。由于这些支配者拥有丰富的数据符号，他们得以将自身的偏好和观念转化为整个场域的正统规则。相对弱势的研究者则常无意识地接受并遵循这些规则，即便这些规则隐含着对其不利的等级秩序。由此可见，符号资本通过对场域规则的影响，可以实现权力结构和学术规范的双向再生产，可以进一步巩固学术权威的等级秩序。换言之，数据符号不仅意味着荣誉和声望，更是一种无形的支配力量，通过对评价体系的塑造持续再生产自身的统治地位。

当数据符号在科学场域内高度集中时，便形成数据符号的垄断，而这种垄断通常通过科研组织结构与评价体系被制度化地体现出来。具有卓越象征资本的精英学者和权威机构通常占据学术组织中的关键位置，比如顶尖大学的终身教授、重要期刊的主编、学术资助机构的评审委员等。这些位置赋予他们正式的决策权，使其能够主导资源分配和评价标准，将个人声望转化为结构性的影响力。在制定规则时，他们倾向于依据自身的成功经验和利益偏好，导致评价体系无意识地偏向已具备较高象征资本的主体。因此，科研评价机制往往奖励符合主流期望、来自知名机构的研究，而对非主流创新或新晋研究者设置更高的准入门槛。这种偏向通过引用率、期刊影响因子等

量化指标显现出来,使象征资本的垄断呈现出“客观化”的表象。强势主体通过将自身优势转化为制度规范,维持对资源与荣誉分配的控制。在科研组织结构中,这体现为少数精英垄断了重要学术职位与荣誉资源,形成封闭的上层圈层,而评价体系的设计又不断强化这种格局,使外部研究者难以突破。这一制度化过程使象征资本的持有者得以长期巩固其学术垄断地位,排他性地掌控学术发展方向与资源配置权。

拉图尔指出,科学知识的生产与传播是在行动者网络中完成的,而“计算中心”因其能够汇集并处理来自各处的数据和信息,在知识传播网络中处于支配地位。这些中心把分散各处的数据资源进行标准化处理,并转化为可计算的符号,通过对远方的事件或现象进行收集、计算和比较,实现“超距离的行动”,从而影响远离中心的知识生产过程。<sup>④</sup>例如,某个大型数据实验室汇聚了全球各地的社会行为数据,不仅在解释社会规律上占据优势,还能够决定哪些问题拥有数据支持,以及哪些研究更值得公开展示。在传播网络中,占据中心地位的节点由此获得垄断式的优势,尤其是关键数据资源与分析工具,主导知识的生产、传播与议题设置。这种网络意义上的垄断与传统象征资本的垄断相辅相成:数据资源的汇聚中心通常也是具有高度象征声望的机构,它们借助网络效应进一步巩固自身在知识生产中的控制权。换言之,拉图尔的理论揭示出传播网络内部的权力不对称性:越是位于网络“计算中心”的行动者,越能够通过信息的集中与再加工,进一步强化对学术资源和话语权的垄断。

从数据资本的获得、数据符号的生成,到传播与奖励机制的运作,科学场域内逐步形成权力高度集中的结构。数据资本必须通过传播体系的制度性验证,才能赢得信任,并借助奖励体系被赋予等级化的象征价值。这些象征资本又通过塑造场域规则而嵌入科研制度,实现垄断性的积累。在此过程中,布尔迪厄所指的象征资本与拉图尔所强调的“计算中心”地位相互呼应,共同作用于学术权力向少数人的集中。最终,他们垄断了计算社会科学的知识生产过程。

## 结论

计算社会科学的兴起,冲击了传统社会科学知识体系。在知识内容方面,传统社会科学研究主要依赖社会调查、统计年鉴、历史与田野资料,关注宏观结构或中观制度,结果一般呈现静态“切片”。实时数字数据使“全时段、全样本”的研究成为可能,进而捕捉微观行为与互动过程,实现跨尺度、多层次的动态观察。与此同时,小数据依然具有独特价值,在回答针对性问题方面不可替代,并且通过与大数据基础设施的结合,正在形成互补性路径。<sup>⑤</sup>这意味着社会科学知识版图并未被单向“替代”,而是在新旧数据形态的交织与张力中不断重组。

在研究范式层面,计算社会科学形成了“计算取向”。传统社会科学研究强调理论假设与经验验证,将理论解释与因果推断视为核心关切,而大数据计算凸显预测和对人类社会某些规律性模式的识别,并能对因果性知识的探究作出新贡献。<sup>⑥</sup>此外,跨学科合作成为常态,经济学、社会学、政治学等与计算机科学深度融合,既拓展了研究议题与方法,也淡化了学科界限。<sup>⑦</sup>因此,计算社会科学不仅是方法工具的叠加,更在根本上推动了学科传统的重新定位。

在这一转向中,数据扮演了核心角色。它既是研究的基础资源,也是决定谁能进入计算社会科学场域的重要门槛。正因如此,本文通过考察计算社会科学研究过程,提出一个基于数据形态的分析框架:将数据分别视为生产要素、资本与符号,并将计算社会科学研究过程划分为进



人、竞逐与垄断三个阶段。该框架可以揭示数据角色变化如何推动知识生产格局的转型。其中，在进入阶段，数据作为新兴生产要素构成知识生产的基本门槛，数据基础设施制造出“准入券”与进入壁垒；在竞逐阶段，数据要素转化为数据资本，各主体围绕数据资源与计算能力展开激烈竞争，通过合作与应用，资本和技术优势逐步集中于少数机构；在垄断阶段，通过传播与奖励，数据资本进一步转化为符号资本，某些优势主体占据中心节点，掌控知识生产过程中的话语权与评价权。

上述分析可以为回答“谁能从事计算社会科学研究”这一问题提供重要线索。尽管计算社会科学在理论上提倡方法与数据的开放共享，但在实际上，知识生产过程却呈现中心—边缘的不平等关系结构。真正能够参与并主导计算社会科学研究的主体，通常可以在基础设施建设、资本积累以及符号权力方面占据结构性优势。首先，数据的基础设施构成计算社会科学研究的硬件门槛。这些设施的获取与维护需要大量的资金、先进技术和人力资源，资源匮乏的机构或个人难以企及。其次，数据本身已演化为战略性资本。领先的科技企业和精英学术机构凭借长期积累的数据和算法优势，形成新进入者难以突破的壁垒。再次，知识生产场域中符号权力的不平等尤为明显。掌握尖端技术和话语权的精英群体更易将其研究成果塑造成“客观真理”，并通过主流期刊、学术会议和评价体系，巩固自身权威地位，排斥其他形式的知识与异质声音。

这些现象表明，原本旨在促进开放共享、广泛吸纳多元主体参与的计算社会科学，在实践中有可能偏离初衷。开放数据与开放方法的倡导，常为数据资源与算力的集中垄断所削弱。新兴的计算社会科学不仅未能消解传统知识生产的等级结构，反而在技术与数据层面强化新的权力不平等格局：掌握数据资源的主体进一步巩固自身优势，而资源匮乏者日益边缘化。

然而，正视这一问题也为重新思考计算社会科学的未来发展提供了重要契机。为重拾开放共享的初衷，未来变革应从以下几个层面展开：在数据层面，构建更为开放的数据协作网络，推动数据共享与资源互通，降低数据垄断形成的准入壁垒；在技术层面，发展公共技术基础设施，包括开放计算平台与工具生态，降低对商业设施的依赖，帮助资源有限的主体获得必要的计算能力；在制度层面，设计并推行资源再分配机制，通过政策和资金支持向弱势或边缘研究群体倾斜，缓解资源配置的不均；在知识治理层面，鼓励不同学科、不同地域与文化背景的知识生产主体积极参与，防止单一范式主导的知识霸权。在数据驱动的人工智能时代，唯有在多层协同发力，计算社会科学才有可能避免优势主体的垄断，真正践行开放、多元和普惠的知识理念。

#### 注释：

①⑤ David Lazer, et al., “Computational Social Science,” *Science*, vol. 323, no. 5915, 2009, pp. 721-723.

② 比如国内有清华大学计算社会科学与国家治理实验室，北京大学政府大数据与公共政策实验室，复旦大学复杂决策分析中心，浙江大学计算社会科学研究中心，南开大学计算社会科学实验室等。

③ 苏毓淞、刘江锐：《计算社会科学与研究范式之争：理论的终结？》，《复旦学报》（社会科学版）2021年第2期。

④ Achim Edelmann, et al., “Computational Social Science and Sociology,” *Annual Review of Sociology*, vol. 46, 2020, pp. 61-81; David Lazer, et al., “Computational Social Science:

Obstacles and Opportunities,” *Science*, vol. 369, no. 6507, 2020, pp. 1060-1062.

⑥ 狄波拉·勒普顿：《数字社会学》，上海：上海人民出版社，2022年。

⑦ Jathan Sadowski, “When Data is Capital: Datafication, Accumulation, and Extraction,” *Big Data & Society*, vol. 6, no. 1, 2019, pp. 1-12.

⑧ 赵一璋：《数字社会的新型资本及其分配特征》，《社会科学》2025年第2期。

⑨ 吕鹏：《计算社会科学：学科体系与领域演进》，《求索》2024年第4期。

⑩ Jake M. Hofman, et al., “Integrating Explanation and

Prediction in Computational Social Science,” *Nature*, vol. 595, no. 7866, 2021, pp. 181-188.

⑪ 王成军：《反思计算社会科学的逻辑：基于拉图尔的“计算中心”概念》，《南京社会科学》2021年第4期。

⑫④④⑥ Bruno Latour, *Science in Action: How to Follow Scientists and Engineers through Society*, Cambridge, MA: Harvard University Press, 1987.

⑬ 吕鹏、范晓光：《计算社会科学导论》，北京：清华大学出版社，2023年。

⑭ 李海舰、赵丽：《数据成为生产要素：特征、机制与价值形态演进》，《上海经济研究》2021年第8期。

⑮ 徐翔、赵墨非：《数据资本与经济增长路径》，《经济研究》2020年第10期。

⑯ 蔡跃洲、马文君：《数据要素对高质量发展影响与数据流动制约》，《数量经济技术经济研究》2021年第3期。

⑰ 皮埃尔·布尔迪厄：《论国家：法兰西西学院课程（1989—1992）》，贾云译，北京：生活·读书·新知三联书店，2023年，第257页。

⑱ 蓝江：《数字资本、一般数据与数字异化——数字资本的政治经济学批判导引》，《华中科技大学学报》（社会科学版）2018年第4期。

⑲ Maryam Farboodi and Laura Veldkamp, “A Model of the Data Economy,” NBER, 2021.

⑳ Lin Cong, Danxia Xie and Longtian Zhang, “Knowledge Accumulation, Privacy, and Growth in a Data Economy,” *Management Science*, vol. 67, no. 10, 2021, pp. 6480-6492.

㉑ 王开科、何强：《基于价值增值视角的数据要素资本化核算研究》，《统计研究》2025年第3期。

㉒③⑩ 谢康、夏正豪、肖静华：《大数据成为现实生产要素的企业实现机制：产品创新视角》，《中国工业经济》2020年第5期。

㉓ OECD, “Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value,” *OECD Digital Economy Papers*, 2013, No.220.

㉔ D. G. Rassier, R. J. Kornfeld and E. H. Strassner, “Treatment of Data in National Accounts,” Paper Prepared for the BEA Advisory Committee, 2019.

㉕③③ 许宪春、张钟文、胡亚茹：《数据资产统计与核算问题研究》，《管理世界》2022年第2期。

㉖⑦⑦ 刘涛雄、戎珂、张亚迪：《数据资本估算及对中国经济增长的贡献——基于数据价值链的视角》，《中国社会科学》2023年第10期。

㉗ 程啸：《论大数据时代的个人数据权利》，《中国社会科学》2018年第3期。

㉘ 李三希、王泰茗、刘小鲁：《数据投资、数据共享与数

据产权分配》，《经济研究》2023年第7期。

㉙④④ 塞德里克·迪朗：《技术封建主义》，陈荣钢译，北京：中国人民大学出版社，2024年。

㉚ 徐翔、赵墨非：《数据资本与经济增长路径》，《经济研究》2020年第10期；张新宝：《论作为新型财产权的数据财产权》，《中国社会科学》2023年第4期。

㉛ 刘涛雄、张亚迪、戎珂、周迪：《数据要素成为中国经济增长新动能的机制探析》，《经济研究》2024年第10期。

㉜ 田利辉、李政、施炳展：《企业上云对其创新的影响：数据要素化的视角》，《世界经济》2025年第1期。

㉝ Simona Abis and Laura Veldkamp, “The Changing Economics of Knowledge Production,” *The Review of Financial Studies*, vol. 37, no. 1, 2024, pp. 89-118.

㉞ 周涛、高馨、罗家德：《社会计算驱动的社会科学研究方法》，《社会学研究》2022年第5期。

㉟ Tania Babina, et al., “Artificial Intelligence, Firm Growth, and Product Innovation,” *Journal of Financial Economics*, vol.151, 2024.

㊱ 皮埃尔·布尔迪厄：《实践理性：关于行为理论》，谭立德译，北京：生活·读书·新知三联书店，2007年，第167—168页。

㊲ 布鲁诺·拉图尔、史蒂夫·伍尔加：《实验室生活：科学事实的建构过程》，上海：华东师范大学出版社，2023年，第218—219页。

㊳ Robert K. Merton, “The Matthew Effect in Science,” *Science*, vol. 159, no. 3810, 1968, pp. 58-63.

㊴ Robert K. Merton, “The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property,” *Isis*, vol. 79, no. 4, 1988, pp. 606-623.

㊵ 皮埃尔·布尔迪厄：《国家精英：名牌大学与群体精神》，杨亚平译，北京：商务印书馆，2018年，第457—459页。

㊶ 苏令银：《大数据时代的小数据会消亡吗》，《探索与争鸣》2019年第7期。

㊷ 唐世平：《计算社会科学科学决策的未来》，《国际政治科学》2023年第3期；孟天广、张静：《基于大数据的实验方法：AI时代实验政治学的前沿推进》，《东南学术》2025年第2期。

㊸ 吕鹏：《计算社会科学方法生态与趋向展望》，《浙江社会科学》2025年第6期。

编辑 杜运泉