

# 创新研究报告

第 48 期（总第 174 期）

2017 年 8 月 22 日

中国科协创新战略研究院

签发人：陈锐

---

## 大数据与社会排斥

**[编者按]** 随着大数据时代的来临，大数据日益被应用到商业和政府的决策之中。大数据在提升经济效率、改善政府治理、提升教育和医疗水平等方面起到不容忽视的积极作用。但技术往往是双刃剑，大数据同时也存在着负面影响，其收集和使用过程中也产生了诸如侵犯隐私、网络安全等众多问题。

围绕以上问题所开展的研究和讨论，更强调已经被采集和使用的数据中存在的问题。然而，即使在大数据成为全球性热点议题的时代，仍有一部分群体的行为并未被纳入大数据的范畴之内，他们的需求被商业和政府机构的决策所忽视，导致其被排斥在各种经济、政治和社会机构之外，以至强化了现有社会的不平等。

鉴于此，本研究围绕大数据社会排斥展开，探讨大数据社会排斥的现状、成因及社会影响。首先，对大数据的定义和运行机制进行简要梳理，分析大数据社会排斥的技术基础；其次，

结合国内外相关领域的前沿研究，阐述大数据社会排斥的现状及其影响；第三，结合欧美等国家的经验，从法律、政府和企业等层面探讨如何改善大数据社会排斥问题；最后，在上述研究的基础上，提出对我国改善大数据社会排斥问题的启发和借鉴。

## 一、大数据：内涵与过程

### （一）内涵

大数据的概念要追溯到 20 世纪 90 年代，指大小超出了传统数据库软件工具的抓取、储存、管理和分析能力的数据集，而且这个数据集在数量、速率与种类上都在不断扩大。换言之，现在数据可以更快获取，有着更大的广度和深度，并且包含了以前做不到的新的观测和度量类型。更确切地说，大数据集是庞大的、多样化的、复杂的、纵深的和分布式的，它主要由各类仪器设备、传感器、网上交易、电子邮件、视频、点击流以及现在与未来所有可以利用的其他数字化信号源产生。

众多大数据的定义中，IBM 从数据体量（volume）、速度（velocity）、多样性（variety）、价值（value）四个维度对大数据所做的界定，得到业界较为一致的认可：

（1）数据体量：表示大数据的数据群体量庞大。数据集的规模不断扩大，已从 GB 到 TB 到 PB 级，甚至开始以 EB 和 ZB 来计数。

（2）速度：数据产生、处理和分析的速度在不断提升，数据流量大。

（3）多样性：表示大数据的类型复杂。旧的数据都可以用二维表结构存储在数据库中，如常用的 Excel 软件所处理的数

据，我们称它为结构化数据。现在，随着互联网多媒体应用技术不断创新，诸多的视频、图片和声音等非结构化数据越来越多，且在数量上与结构化数据相比都是有着云泥之别的。

(4) 价值：随着大数据的体量不断提升，单位数据的价值密度在不断下降，然而数据的整体价值在不断提升。

从以上“4V”的维度可以看出传统大数据与大数据的区别，如表 1 所示。

表 1 传统数据与大数据的区别

维度	传统数据	大数据
数据体量	GB, TB	TB—PB 级以上
速度	数据量相对稳定，增长缓慢	持续、实时产生数据，增长速度较快
多样性	主要以结构化数据为主，缺乏数据源	结构化、半结构化、音频视频、多维多源数据
价值	统计和报表	分析预测、数据挖掘、决策

## (二) 过程

数据从“小”数据变成“大”数据，主要经过采集、汇编和整合、分析以及使用四个阶段。在采集阶段，并非所有的数据都被称为“大数据”。商业机构从各种各样的来源采集数据，例如，追踪消费者浏览网页或网上购物的活动信息，以及在移动环境中追踪并将消费者的活动链接到应用程序中采集用户习惯和偏好信息。从更广泛的方面来说，跨设备跟踪提供了在台式机、笔记本、平板电脑、可穿戴设备和智能手机上与同一消费者进行互动的能力，同时线上和线下的信息均可使用。最后，数据采集也会发生在离线状态，例如，通过保修卡、调查、抽奖活动，甚至信用卡购物等状态。

二是汇编和整合阶段。汇编数据的商业实体包括在线广告网络、社交媒体企业、大型银行和零售商。该汇编和整合数据的商业实体的一个重要类别是数据代理商，它们综合不同来源的数据，建立消费者的个人资料库。

三是数据分析阶段。分析的一种形式是描述——目标是揭示并总结数据集中存在的模式或特性。预测性数据分析指利用统计模型生成新数据。在市场环境下，大数据分析的目的是对消费者的选择做出推断。对消费者数据的分析，往往对企业和消费者有价值，因为它可以引导新产品和新服务的开发，预测个人喜好，有助于为消费者提供定制服务的机会，引导个性化营销。与此同时，倡导者、学者及其他人提出要关注大数据分析的某些应用是否会损害消费者利益，特别是低收入和服务匮乏人群。

四是使用阶段。以往的一些研究主要聚焦于大数据采集、融合和分析三个阶段，主要探讨信息再分析利用的过程。本研究重点介绍使用大数据带来的收益和风险，它可以提供很多改善社会服务的机会。除了更有效地为消费者匹配产品和服务，大数据还可以为低收入和服务匮乏社群提供教育、信贷、医疗和就业机会。

## **二、大数据社会排斥的现状、成因及影响**

### **（一）大数据社会排斥现状**

社会排斥概念源于法国，后被欧洲国家广泛采用，应用于诸如教育学、社会学、心理学、政治学以及经济学等领域的研究。尽管不同学科对如何定义社会排斥存在一定分歧，但业已形成一定共识，即社会排斥指某些群体被系统化地排斥在就业、政治参与、住房、教育、医疗等机会之外，被剥夺获取这些资

源的权利，进而形成了社会的撕裂。

近年来，大数据逐步成为炙手可热的话题，受到广泛的关注。随着智能手机以及其他智能设备的广泛应用，移动应用程序（APP）在生活中占据日益重要的地位以及物联网的兴起，无数的个体产生着大量的数据。与此同时，数据采集、存储与分析的成本急剧下降，以及数据代理商整合线上和线下能力的增强，为企业收集近乎无限量的消费者信息以及进行永久性存储提供了便利。

然而，在收集、存储和处理大规模信息的过程中，存在着社会排斥的问题，即包括低收入群体在内的底层弱势群体被排斥在大数据带来的各种机遇和资源之外。美国联邦贸易委员会（Federal Trade Commission, FTC）于2014年发布了数据代理行业的研究报告，报告通过对9个数据代理商的研究发现，数据代理商基于不同来源的线上和线下数据，结合一定的算法，来推算不同种族、收入、宗教、年龄和健康状况群体的消费行为。然而，基于此类算法进行的消费群体划分，不仅有可能将一部分没有消费能力的群体排斥在外，而且极有可能在商业逻辑下形成排斥的合法化，从而加剧现有的社会不平等。

## （二）大数据社会排斥的类型及成因

大数据社会排斥分为两种形式：数据收集层面的排斥和使用层面的排斥。大数据收集层面的排斥，指有些群体，尤其是弱势群体，其行为信息未能被大数据收集系统所采集，导致其被排斥在基于大数据的政府和商业决策依据之外。而使用层面的排斥，指有些弱势群体，即使其使用智能设备和应用，其行为产生了一定的数据并且被大数据采集系统所收集，但以商业

机构为主的相关机构由于商业利益等原因，在决策过程中对这些群体的信息进行过滤，将其排斥在外。

对于采集层面的排斥，乔纳斯·莱尔曼的研究中所列举的两个假想案例，是此问题的真实写照。第一个案例主角是生活在美国曼哈顿的30岁女性白领工作者，她过着现代白领的典型生活：拥有智能手机，使用包括Facebook在内的各种主流社交媒体，经常旅行并且将定位和图片等相关信息分享到社交媒体平台，钱包中装有多种银行卡和超市、商场等机构的积分卡，车上配备导航系统。这位女主人公以及类似她的众多个体，每天的行为所产生的大量消费、交往、线上活动以及空间移动的数据，大部分为政府机构和商业公司所收集，成为政府和商业机构决策的重要依据。

相反，第二个案例的主角则处于截然不同的境遇：他生活在美国最贫穷的城市卡姆登（Camden），距曼哈顿仅2小时车程；他没有正式工作，在一个餐馆兼职，有限的收入以现金的方式私下支付；他没有手机、电脑，也无法接入互联网；他很少旅行，也没有护照、汽车或者全球定位系统；他有时会上网，但也仅限于在当地图书馆；他会使用公共交通，但一般使用现金支付车费。他的生活方式决定着其很少参与到产生数据的行为，因此他和他所代表的群体很难生产出能被收集的数据。

莱尔曼提出，这两位主人公的不同境遇正体现了大数据的社会排斥问题。两个主人公，由于居住在发展程度有巨大差距的两个城市，处于不同的社会阶层，过着截然不同的生活方式，而这些不同导致前者的行为产生了大量数据而且被政府和商业机构所收集，后者的行为很难产生出数据，更不用说被相关机

构所收集。因此，后者所代表的群体被排斥在大数据以及基于大数据分析所作的决策之外。

以上讨论的采集层面的排斥，其根源在于缺少接入智能手机和网络以及使用能生产出大数据的电子设备及应用。此类排斥在互联网和智能设备扩散初期较为显著，主要表现为技术采纳者和未采纳者之间的鸿沟。在发达国家，随着智能设备在全社会的扩散，大数据采集层面的排斥问题较容易解决。如皮尤报告表明，发达国家智能手机普及率已经达到较高水平，韩国为 88%，美国为 72%，英国为 68%，德国 60%，意大利为 60%。然而，在发展中国家和经济较为落后国家，智能手机普及率仍低于全球水平，如非洲地区很多国家的普及率低于 20%。

然而，通过提高智能设备和应用的普及率来解决数据采集层面的排斥问题，并非意味着大数据社会排斥问题的完全解决，因为尚存在使用层面的排斥问题。具体而言，即使同样有机会拥有智能设备、使用各种智能应用并产生一定数量的数据，但并非所有数据都能被商业之类的机构认为是“合格”的数据。在消费社会，不具备消费能力的弱势群体被视为“不合格的消费者”，从而被边缘化和排斥。因此在大数据系统中，这些“不合格的消费者”的行为所产生的数据，更倾向于被认定为“不合格”而被排斥在外，从而被剥夺享有社会和商业利好的机会。

在美国联邦贸易委员会组织的大数据排斥问题的座谈会中，商界和学术界人士均对大数据使用层面所产生的排斥提出了意见，包括大数据准确性、完整性和代表性等维度在内可能会出现的质量问题，以及潜在消费数据中出现的未修正的偏差问题。

例如，有与会学者指出，数据收集、分析和过程中所隐含的偏差，如果不加以修正，有可能导致某些群体被排斥在商家的消费决策之外，使其不能公平获取市场机会，从而强化甚至加剧现有的社会不平等。

### **（三）大数据社会排斥的影响**

#### **1. 经济层面**

在经济层面，被大数据排斥在外的群体将会面临被剥夺某些经济权利和机会的状况。随着大数据时代的到来，商业机构日益将大数据纳入其决策体系，其相关运营业务都将依据数据挖掘的结果展开，而未被纳入其数据系统的社会群体的需求将会被忽视，也很难享有商业机构的各种促销优惠政策。更进一步来说，商业机构极有可能根据其数据分析结果决定是否在相应区域开设分店或者分支机构，一旦某些群体信息未被收集或者虽被收集但被认为无效，那么他们则失去了日常消费的便利性和相应的就业机会。

在商业机构的大数据分析中，基于一定算法的决策决定了面向消费者的商品的选择，价格的制定，以及为消费者提供服务水平的设定，而在此过程中，如果不加以干预，极有可能导致现有社会优势群体享有更专业化和高质量的服务，而被排斥在外的群体则也被排斥在这些产品和服务之外。

#### **2. 社会层面**

在社会层面，大数据所产生的社会排斥问题，不仅反映了当今社会不平等的现实，同时又通过其排斥进一步强化甚至加剧了现有的社会不平等。哈佛大学斯威尼教授对不同种族的名字在搜索引擎的搜索结果的差异进行研究，揭示了大数据情境下的种族排斥问题。斯威尼教授通过在谷歌等搜索平台分别

搜索具有不同种族特征的名字，并对搜索结果进行比较，发现黑人名字的搜索结果中具有犯罪记录人的百分比，要比白人高25%。也就是说，如果在搜索引擎中搜索黑人名字，所推荐的结果更容易和罪犯有关联。这种印象将恶化黑人群体在就业市场的竞争地位，强化黑人现有的弱势地位，进一步加剧社会的不平等问题。

### 3. 政治层面

被大数据所排斥的群体在争取政治权益上处于严重弱势地位。随着越来越多的政府机构对大数据日益重视，大数据在政府决策中起着日益重要的作用。那些优势群体，由于其在产生数据的数量和质量上所占的优势，其声音更容易被政治机构所听到，其诉求更易被满足。反之，被排斥在大数据之外的群体，其发声的渠道和机会将会严重受限，其利益和诉求更难以被满足。在资源有限的情况下，如果分配向既有优势群体倾斜，那么势必会加剧现有社会的不平等。

## 三、国外相关干预措施

### 1. 法律层面：加大对大数据体系的监管

面临大数据所引发的排斥问题，美国和欧洲均采取了法律层面的监管。美国通过识别现有法律中哪些适用于大数据相关业务，运用现有法律对大数据实践进行规范。如美国联邦贸易委员会于2014年判定网站 Instant Checkmate 因提供背景调查作为筛选租户和雇员的工具，而违反了美国保护个人信用信息的《公平信用报告法》（*Fair Credit Reporting Act*），对其处以525000美元的罚款，并对该公司实施了永久禁令。

同样，欧洲也出台了相关的条例，试图解决大数据所带来的排斥等众多问题。2016年4月，欧盟议会出台了《一般数据保

护条例》（*General Data Protection Regulation, GDPR*），对收集、存储和使用个人数据进行了严格规范，这是近 20 年间欧盟第一次出台类似条例。该条例强调的重点之一是对大数据收集、存储和使用中可能存在的歧视行为作出明确规定，以确保某些弱势群体不被排斥在外。

## 2. 企业层面：引导企业行为以避免大数据排斥

美国联邦贸易委员会前会长拉米雷斯在 2014 年举办的以大数据与社会融合与排斥为主题的研讨会中提出，为了避免大数据所引发的严重社会影响，需要企业在伦理和业务层面引起关注。在伦理层面，不仅需要促使企业充分认识并重视大数据对低收入和弱势群体的负面影响，提升企业在大数据收集和使用过程中的社会责任意识，而且，也应鼓励企业在设计其分析体系、算法和预测产品中防范偏见。

以美国波士顿市为例，为了及时发现道路问题并改善道路质量，波士顿设计发布了名为“Street Bump”的手机应用。此项应用采纳了大数据的理念，召集市民志愿者下载此应用并在行车时使用，以收集波士顿市的整体路况数据，从而识别需要修整的路段并加以修缮，以及规划长期投资。该应用已获得广泛关注，并被列入美国大数据报告中的成功案例。其成功之处，除了大数据思维的应用，更重要的是应用的设计中充分考虑了大数据的偏差及其社会排斥问题。

具体而言，应用的开发机构考虑到低收入群体较之富裕群体更少使用智能手机，有可能导致此应用收集的数据主要为富裕社区的数据，如根据此数据规划道路修缮和投资，将导致公共资源的投入集中向富裕阶层倾斜，从而加剧贫困与富裕社区的不平等问题。由于预先意识到以上问题，其数据收集并非完

全依赖基于志愿者收集的大数据，而是为各个社区的公路巡路员配备此应用，发动巡视员检查并上报负责区域的路况，作为收集的公众数据的补充。由于事先充分考虑并重视大数据可能产生的社会排斥问题，并作出了相应合理的调整，此项应用既发挥了大数据在公共服务中的优势，也避免了社会排斥问题的出现。

### **3. 政府层面：积极引导大数据发挥积极作用，为弱势群体服务**

政府作为大数据的收集和使用，在直接使用大数据进行治理决策的过程中，也需要开发大数据中有益于弱势群体的一面，充分发挥大数据的积极作用。以美国的纽约市为例，每年有大概 200000 户租户被房东驱逐，处于无家可归的境地。虽然政府的社会福利机构一直致力于帮助这一群体，但如何及时、有效识别哪些人面临无家可归的风险，仍存在一定难度。为了解决此问题，非营利数据分析公司 SumAll 建立了综合法庭记录、庇护史（shelter history）和人口信息的数据库，开发了数据分析工具，通过相应算法识别哪些人面临无家可归的风险，使得福利机构能有针对性地为困难群体提供帮助，避免其流落街头。

## **四、启示和借鉴**

### **1. 充分认清大数据采集和使用层面的排斥问题**

大数据的排斥，不仅包括信息采集层面的排斥，也包括虽然信息被收集但因被认为“不合格”而被排斥。对于采集层面的排斥，需要解决的是智能设备和应用的普及问题，使更多的群体尤其是弱势群体能接入智能产品，确保其信息被合理采集，提升大数据的准确性、完整性和代表性，其行为信息被纳入商

业和政府机构决策的参考范畴，解决信息采集层面存在的排斥问题。

而信息使用层面的排斥，即信息是否被使用的问题，决定着排斥问题是否能从根本上得到解决。具体来说，即使所有弱势群体得以采用智能设备和应用，产生了相应的行为数据，并进入了数据收集的系统中，但如果相关机构在使用过程中，将弱势群体所产生的数据认为是“无价值”的数据，仍根据优势群体的数据信息作出相应决策，那么本质上来说弱势群体仍然被排斥在外，而且会被进一步边缘化，社会撕裂问题并未得到解决。

## 2. 通过法律法规规范企业大数据使用

在大数据收集、储存和使用系统中，商业机构由于其逐利的本质，为了实现商业利益的最大化，更易在商业逻辑的掩盖下，树立其排斥行为的合法性。

因此，首先需要加强对涉及的企业、数据代理公司等各个环节所涉及的机构加强企业责任感和伦理层面的引导，鼓励其决策过程中充分考虑到大数据有可能产生的社会排斥问题，并采取相应的行为以减少排斥问题的产生。

其次，除了伦理层面的建设，也需要通过实施相应的法律法规。在规范层面，需要审视现有的相关法律法规，是否适用于大数据采集、存储和使用过程，将大数据监管纳入现有的体系。除了沿用现有法律法规，考虑到大数据相关环节存在的一些特殊性，因此需要根据新情况、新问题，制定新的规范，以实现有更有针对性的监管。

## 3. 政府使用中发挥大数据的积极作用

除了通过法律法规来规范商业机构的行为，减少其基于大

数据的商业决策中对弱势群体排斥的行为，政府自身决策过程中，也需要事先充分考虑到大数据的排斥问题，以进行相关的规划和事后干预。如上文提到的波士顿政府基于大数据规划道路投资和维护过程中，由于充分考虑了大数据社会排斥的情况，对弱势群体所在的社区进行了补充性措施，使得政府决策充分考虑到了全体公民的意愿和需求。除此之外，政府也需要发挥联结作用，将商业界、学术界和政府机构等不同领域与大数据收集、存诸和使用有关的人士联合起来，定期以研讨会等形式，充分听取业内人士意见，共商如何解决大数据的社会排斥问题。

## 参考文献

- [1] BAUMANZ. Work, consumerism and the new poor[M]. Society, 2013. UK: McGraw-Hill Education.
- [2] EWINGM. Big data could help some of the 200000 NYC households that get eviction notices this year[EB/OL]. NextCity. 2015. <https://nextcity.org/daily/entry/evicted-help-new-york-homelessness-prevention>.
- [3] EINAV L, LEVIN J. The data revolution and economic analysis[J]. 2014. Innovation Policy and the Economy, 14:1-24.
- [4] RAMIREZE, BRILL J, KOHLHAUSEN M. Big data: a tool for inclusion or exclusion[J]. 2016. Understanding the Issues: 8-11, 25-27.
- [5] FTC. Data BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY [EB/OL]. [www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountabilityreport-federal-trade-commission-may-2014/](http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountabilityreport-federal-trade-commission-may-2014/)

140527databrokerreport.pdf.

[6] FTC. Two data brokers settle ftc charges that they sold consumer data without complying with protections required under the Fair Credit Reporting Act[EB/OL]. <http://www.ftc.gov/news-events/press-releases/2014/04/two-data-brokers-settle-ftc-charges-they-sold-consumer-data>.

[7] GOODMAN B, FLAXMANS. European Union regulations on algorithmic decision-making and a “right to explanation”[Z]. 2016. arXiv preprint arXiv:1606.08813.

[8] Hwang M. FTC Report’s Impact May Not Be Limited To Data Brokers[J]. 2013.

[9] LERMAN J. Big data and its exclusions[J]. 2013. STANFORD LAW REVIEW ONLINE, 66:55–63.

[10] POUHTER J. Smartphone ownership and Internet usage continues to climb in emerging economies[R]. 2016. Pew Research Center.

[11] Ramirez E. A tool for inclusion or exclusion[Z]. 2014. Opening Remarks of Chairwoman Edith Ramirez, FTC Big Data Workshop.

[12] SNIJDERS C, MATZAT U, REIPS U D. “BigData” :big gaps of knowledge in the field of internet science[J]. International Journal of Internet Science, 2012, 7(1) :1–5.

[13] SHMUELI G, BURKOM H. Statistical challenges facing early outbreak detection in biosurveillance[J]. 2010. Technometrics, 52(1) :39–51.

[14] SWEENEY L. Discrimination in online Ad Delivery 56 COMMUNICATIONS OF THE ASS’ N FOR COMPUTING MACHINERY 44[Z/OL]. <http://dataprivacylab.org/projects/onlineads/1071-1.pdf>.

[15] WU X, LI Y, ZHANG Q, et al. Knowledge engineering with big data[J].

2015. IEEE Intelligent Systems, 30(5):46-55.

[16] Young M. A Human-Centered Approach to Data Privacy: Political Economy, Power, and Collective Data Subjects[J]. 2016.

(作者: 方伟 王达)

## 文章来源

《2017年中国社会学年会——科学社会学论文集》



创新研究公众号



中国科协创新战略  
研究院官网

编辑: 高晓巍

---

联系电话: 13811520165

电子邮箱: [gaoxiaowei@cast.org.cn](mailto:gaoxiaowei@cast.org.cn)