

一个数据要素的经济学新理论框架

蔡思航，翁 翕

（北京大学 光华管理学院，北京 100871）

摘要：本文基于一个数据要素的经济学新理论框架，围绕平台、流量和数据等数字经济新元素，研究了数据要素市场发展中的经济规律和关键问题。该框架主要基于博弈论与信息经济学，强调数据要素的非竞争性和副产品特性是其区别于其他传统生产要素的核心特点和优势，规模收益递增和网络效应意味着数据要素在价值创造和分配中可以实现“1+1>2”的效果，通过放大、叠加、倍增作用提高其他要素的全要素生产率。本文认为，当前数据要素市场的发展仍需解决如何合理衡量数据要素收益和贡献，以及构建初次分配、再分配和第三次分配机制等关键问题，应鼓励市场主体逐步探索和完善数据交易模式和定价体系，并通过税收设计和共享开放实现数据红利的再分配，从而更充分地提升数据要素价值。

关键词：数据要素；数据要素市场；信息经济学；价值创造与分配

中图分类号：F49 **文献标识码：**A **文章编号：**1000-176X(2024)05-0033-16

一、引言

从新一轮科技革命和产业变革的大趋势来看，第四次工业革命以数字化、智能化、网络化为特点，而数据资源作为重要的生产要素，被认为是21世纪的“黄金”“石油”。党的十九届四中全会将数据作为五大生产要素之一。2020年，中共中央、国务院发布的《关于构建更加完善的要素市场化配置体制机制的意见》《关于新时代加快完善社会主义市场经济体制的意见》明确提出，培育和发展数据要素市场。《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》进一步提出，建立数据资源产权、交易流通、跨境传输和安全保护等基础制度和标准规范，推动数据资源开发利用。2022年12月，《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》（简称“数据二十条”）发布。2023年，国家数据局组建。2024年1月，《“数据要素×”三年行动计划（2024—2026年）》发布。这些文件和制度安排表明，从顶层设计角度，党中央、国务院将建设数据基础制度、整合共享和开发利用数据资源、发展和完善数据要素市场等置于中国数字经济发展战略的重要位置。

收稿日期：2024-03-02

基金项目：国家自然科学基金重大项目“数字经济中数据要素有效使用与消费者保护”（72192843）；国家自然科学基金杰出青年科学基金项目“信息经济学理论”（72225001）

作者简介：蔡思航（1993-），女，江苏南通人，博士研究生，主要从事产业经济学和信息经济学研究。E-mail: scai@stu.pku.edu.cn

翁 翕（通讯作者）（1983-），男，江西景德镇人，教授，博士，博士生导师，主要从事博弈论、信息经济学和组织经济学研究。E-mail: wengxi125@gsm.pku.edu.cn

将数据确立为重要生产要素是中国在全世界范围内的首创,这也意味着现有的经济学理论并没有系统地数据要素进行分析。为了弥补这一理论空缺,本文从数字经济商业模式的核心元素出发,提出了一个涵盖数据要素开发、流通、应用和分配的经济学新理论框架,结合数据要素特性梳理了数据要素市场发展的关键问题,并针对数据要素的价值创造和分配提出政策建议。本文提出的数据要素的经济学新理论框架有别于新古典经济学中消费者和厂商的两分法,新增了平台和机器两个参与者,同时着重强调数字经济中流量、数据、创新和个性化服务这四大元素。数据要素的经济学新理论框架可以概括为一个正反馈循环:在数字经济平台上,消费者通过浏览和购买等行为产生数据;机器对数据进行开发加工和建模分析后,可以赋能厂商进行产品和服务创新;平台通过为消费者匹配个性化服务,吸引新的流量和消费行为,促进更多数据的产生。在这个过程中,平台不仅可以直接通过流量变现获取收益,还可以借助数据更加高效地利用流量。其中,平台、消费者和厂商对于各自的定价、生产及购买的决策,可以通过构建不同形式的博弈问题来更好地分析和理解。平台技术、消费者偏好和厂商成本等通常为私有信息,适合用信息经济学理论来研究各类数据流通交易和共享开放机制的结果及影响。

在基于博弈论和信息经济学的数据要素的经济学新理论框架下,数据要素的非竞争性和副产品特性是其与其他传统生产要素的核心区别,在价值创造方面赋予了数据要素相较于其他生产要素无可比拟的优势。数据要素的规模收益递增和网络效应特性意味着其在价值创造方面可以实现“1+1>2”的效果,因而数据要素应与实体经济深度融合,赋能传统产业转型升级,催生新业态新模式,通过数据要素的放大、叠加和倍增作用提高其他要素的全要素生产率。伴随着互联网经济的发展和大数据时代的到来,数据要素市场具有“类基础设施”的作用。在海量数据基础上,通过大数据分析、云计算、人工智能和区块链等技术,数据要素可以更好地帮助劳动力、土地、资本、技术等传统要素实现有效配置和价值提升。数据要素还天然地具有三次分配属性,因而可通过数据共享和开放等方式在促进数据流动的同时实现数据红利的更公平分配。

数据要素的经济学新理论框架可以帮助我们探讨关于数据要素价值创造与分配的重要问题。在数据要素价值创造方面:首先,在信息价值理论的指导下,数据要素贡献衡量可以结合博弈论中的Shapley值理论以公平准确地计算数据的贡献值;其次,数据要素交易模式需结合数据的非竞争性和网络效应等特性,在机制设计、供需匹配等理论指导下,构建多元化交易市场,力求实现数据要素的有效利用;再次,数据要素定价需要以激励相容为前提,促使买卖双方依照真实的价值或成本评估参与交易,减少套利空间和欺诈行为;最后,在差分隐私和联邦学习等隐私计算技术基础上,用算法博弈论的工具研究数据要素确权与隐私保护等问题,妥善解决数据要素价值挖掘与隐私安全保护之间的矛盾。在数据要素分配方面:首先,垄断和寡头竞争以及平台多边市场理论有助于厘清数据要素在不同市场主体之间初次分配的份额;其次,数据要素的再分配可以结合初次分配的份额和数据要素的特性,在最优税收理论的指导下,探索数字税或利润税改革;最后,数据要素的三次分配应借鉴公共信息披露理论,解决什么是数据最优开放共享模式,数据应对哪些用途场景和需求方免费等核心问题。

二、文献述评

本文提出的数据要素的经济学新理论框架涉及数字经济中的流量、数据、创新和个性化服务四大元素,现有研究虽然对这四个元素都有所涉及,但缺乏从数字经济视角的深入研究。现有的少量研究大多从数字经济视角出发,仅孤立地讨论某一元素,而没有把四个元素整合为一个有机整体,主要体现为以下四个研究方向:

第一,流量的重要性背后反映了一种特殊资源的稀缺性,这种特殊资源即消费者的时间或注意力。传统经济学理论假定人的注意力是有限的,可以关注到所有与决策有关的因素。但数字经

济中的商业实践以及越来越多的实证研究表明,决策者的注意力往往有限甚至是稀缺的,因而需要在消费者理论和竞争性均衡等微观经济学框架中作出相应调整,如De Clippel等^[1]和Gabaix^[2]关注的理性疏忽(Rational Inattention)模型等。但现有文献鲜有将理性疏忽模型应用于数字经济场景进行研究,特别是探讨数字经济平台如何获取流量、分配流量等。

第二,随着数字经济的蓬勃发展,数据要素的经济学理论得到了越来越多经济学家的关注。国内外学者对数据要素的特性进行了理论探索。同时,学者们也对现实中数据要素的开发、流通、应用和分配过程中的实践经验进行吸收和总结,不断为理论研究开拓新方向。数据要素具有一定的非竞争性(Non-Rival)且可以无限复制、重复使用^[3],导致数据资产无法直接套用传统的机制设计理论来构建其交易机制。传统的机制设计理论主要考虑竞争性私有物品的交易。以拍卖为例,只有最终的赢家可以获得被拍卖品,而其他买家均一无所获。但是对于数据资产而言,一个机构提供的数据能够被多个买家同时运用于模型的训练以及预测。在这种情况下,采用“出价最高者得”的拍卖机制来交易数据显然不是最有效的资源配置方式,对企业而言也无法实现利润最大化^[4]。因此,一个好的数据交易机制需要设计出卖方提供什么样的数据给不同的买家,以及从不同买家处索取什么样的价格^[5]。

第三,虽然已有很多文献关注创新,但鲜有研究关注在数字经济背景下的创新。在数字经济背景下,一些学者正在探索数字平台垄断对行业内企业创新的影响,但现有文献得到的结论不尽相同。一部分学者认为,垄断可能抑制数字企业的创新。Kamepalli等^[6]从初创企业估值的视角出发,指出垄断平台对初创企业的收购行为会导致风险资本家更不愿投资初创企业,加剧初创企业融资难问题,不利于行业整体的创新。其实证研究也表明,Facebook和Google等平台对相同或邻近行业的初创企业进行的大规模收购减少了初创企业获得的投资。另外,由于数字平台中存在网络效应,大型数字企业有动机进行猎杀式并购(Killer Acquisition),即并购后关停被收购企业的所有服务。如果缺乏有效的管制,会导致数字经济领域猎杀式并购盛行,将对初创数字企业的创新活动产生负面影响^[7]。不过,也有一部分学者认为,数字行业市场集中度的提高会对数字企业的创新产生正向影响。Gautier和Lamesch^[8]认为,由于数字企业的创新成果难以被单独交易,因而数字企业的兼并是较好的技术转移渠道。如果数字企业的兼并活动较为活跃,初创数字企业将基于对技术转移的良好预期更有动力进行创新。

第四,在个性化服务方面,现有文献主要探究个性化定价问题,即平台在获得消费者数据后,可以利用数据推测消费者的支付意愿,据此通过价格歧视最大限度榨取消费者剩余以提高利润。现有基于价格歧视理论的研究^[9-11]认为,个性化定价会导致剩余分配从消费者向生产者倾斜,从而损害消费者利益。本文认为,在个性化定价之外,学者还应关注更好地满足消费者长尾需求的个性化服务。

三、基于数据要素的数字经济商业模式

数字经济发展催生出有赖于流量、数据等新型资源或生产要素的商业模式。但传统新古典经济学理论分析局限于劳动、资本、土地等传统生产要素,已无法适用于研究与新型生产要素相关的经济规律。例如,Samuelson和Nordhaus^[12]所提出的传统经济学基本分析框架没有融入平台、流量和数据等数字经济新元素。同时,传统经济学理论通常基于规模报酬不变、完全信息、完全竞争市场和无外部性等假设,并要求生产要素和产品具有竞争性、排他性和边际成本递增等特性,这些假设均不符合数据要素特性。因此,Samuelson和Nordhaus^[12]提出了经济学应该关注生产什么、如何生产和为谁生产这三个核心问题,而新古典经济学理论实际上主要关注的只是如何生产的问题,而生产什么和为谁生产这两个问题都通过完全信息假设被解决了。传统经济学基本分析框架^[12] 29如图1所示。

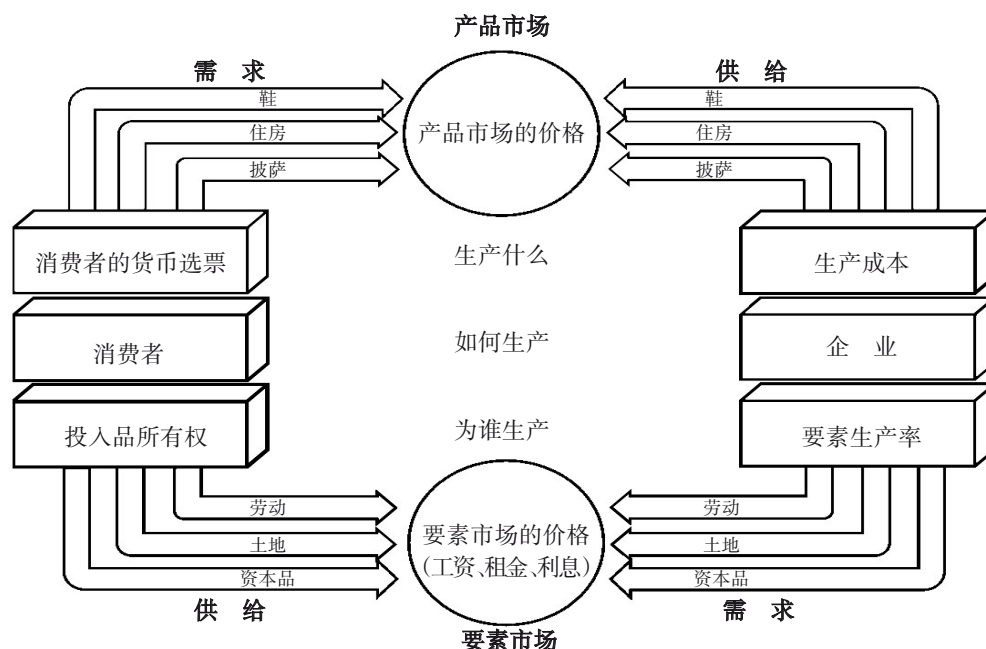


图1 传统经济学基本分析框架

数据要素的经济学新理论框架将数字经济商业模式概括为流量、数据、创新和个性化服务的循环，其中，数据是带动这个循环的核心生产要素。基于此，本文绘制了基于数据要素的数字经济商业模式图，具体如图2所示。由图2可知，在数字经济平台上，消费者的浏览和购买等行为产生原始数据，大量数据经过开发加工，通过机器学习等方法建模，可以赋能厂商更好地把握市场趋势并开展创新，为消费者提供丰富的个性化产品和服务。同时，平台可以提高商户和消费者的匹配效率，为个性化服务吸引新流量，进而产生更多的数据，并形成数字经济的正反馈循环。

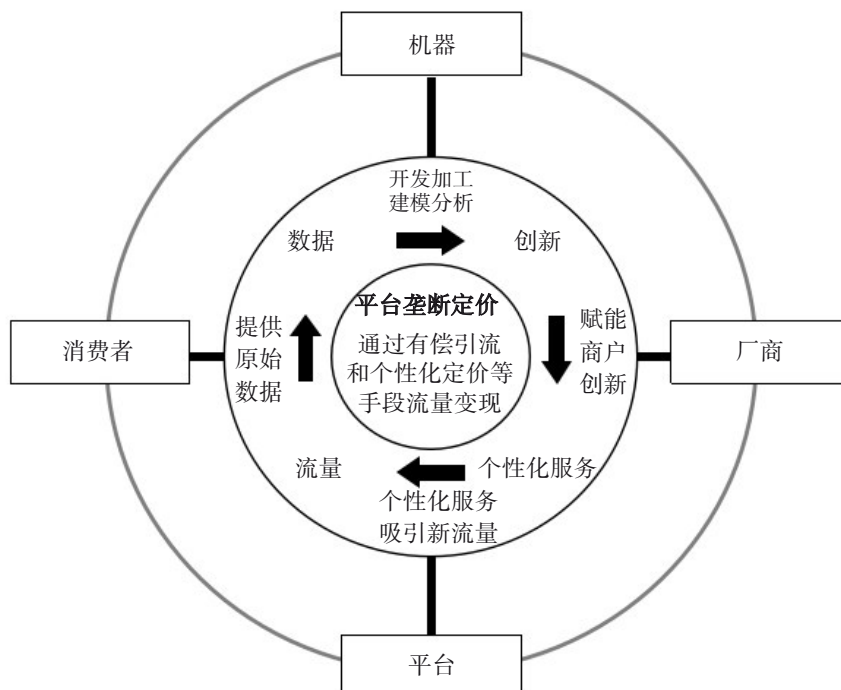


图2 基于数据要素的数字经济商业模式图

(一) 流量是数字经济的基础

流量通常指互联网平台在一段时间内用户的访问量,一般用每日活跃用户数、月度活跃用户数衡量。数字经济与传统经济的主要区别在于后者对流量的空前重视。在传统经济中,如果消费者仅仅访问线下商铺而没有发生购买行为,其对企业利润并没有任何贡献。但在数字经济中,交叉补贴的商业模式使平台可以直接通过流量变现获取收益,而消费者对平台的访问也可以转化为数据,帮助平台和厂商优化商业决策。

流量在中国数字经济的高速发展中发挥了重要作用,特别是2007年以后,移动互联网带来了海量的用户群,大规模的用户增量为数字平台发展带来了新的契机。第45次《中国互联网络发展状况统计报告》显示,2019年底,中国互联网普及率达64.54%,其中,手机网民规模达8.97亿人,占比99.26%^[13]。基于超大规模的人口优势,中国头部平台企业迅速发展壮大,使中国成为仅次于美国的平台经济大国。根据中国信息通信研究院政策与经济研究所数据,截至2019年底,中国市值超10亿美元的平台企业增长至193家,平台总价值达到2.35万亿美元,腾讯和阿里巴巴均跻身于2019年全球市值排名前十位公司之列^[14]。但由于消费者注意力的稀缺性导致了流量的有限性,流量在经历了快速增长后逐渐趋于平缓。截至2021年12月底,中国手机网民规模达10.29亿人,2021年共增加4373万人,同比增速放缓^[15]。2022年第一季度,国内用户平均安装APP的数量为66个,2021年第二季度、第三季度和第四季度分别为66个、66个和65个。这意味着用户每安装一个APP就会删除一个APP以维持动态平衡^[16]。可见,中国流量增长动力不足,互联网环境已不是任由开发的“蓝海”,平台的增长模式需要进行根本性重构。

在互联网“流量见顶”的压力下,各类平台为了争抢流量引发获客成本高企。流量在数字经济中的特殊作用使平台企业之间的流量竞争越来越激烈。例如,各个平台纷纷开启直播带货,甚至为吸引用户注意力,不惜采用过度宣传甚至低俗宣传的方式。同时,这种流量的竞争也反映在日趋提高的获客成本上面。以市场及销售费用/年度活跃买家的方式测算,拼多多和阿里巴巴旗下电商平台的获客成本分别从2017年的5元/人和43元/人,增加至2020年的52元/人和81元/人^[17]。日益高企的获客成本被平台转嫁给平台内的中小企业,从而削弱了数字经济对实体经济的支持作用,并且限制了数字经济的未来发展空间。

(二) 数据是数字经济的核心

如前文所述,互联网流量成本日益高企意味着平台需要更有效地利用流量,而数据则是流量高效使用的关键“钥匙”。数据来源于经济活动中的各类主体对现实世界的数字化记录。数字经济的快速发展引发了数据的爆发式增长和海量集聚,使数据成为重要的战略资源和新型生产要素。作为一种新型生产要素,数据具有搜索成本低、复制成本为零、传输成本低、非排他性等特征,能够多场景应用、多主体复用,可以通过优化资源配置,提高劳动和资本等其他要素的投入产出效率,实现对经济增长的乘数效应。

此外,数据规模报酬递增使得数字经济具有网络效应、规模经济和范围经济等特点^[18-19]。一方面,少量或低质量数据可能价值不高,只有聚合的大量数据或有标注的高质量数据才有实用价值。另一方面,在数据积累的初期,新企业和中小企业可能成长速度较慢,而头部企业则可以获得快速增长,从而导致企业规模两极分化,不利于市场生态与竞争秩序的形成与健康发展。

数据在现实中已被广泛用于训练人工智能模型,以改善各类决策和行为。2023年,生成式人工智能和大模型的快速发展离不开丰富、高质量和多元化的数据支持。得益于大量数据的支持,生成式人工智能已逐渐能够生成高质量的文本、逼真的图像和视频等。随着数据规模的不断扩大和数据质量的提高,人工智能技术将为各个行业创造更大的价值。得益于人工智能技术的发展,平台经济领域中元宇宙、自动驾驶和数字人直播等新产业、新业态、新模式出现爆发式增长,也成为推动中国经济高质量发展的新质生产力。以电商平台为例,通过运用云计算、人工智

能、大模型等数字技术, 电商平台得以整合产业链资源、优化价值链, 降低了企业运营成本、提高了企业运营质量和效率、按需调整了生产活动、实现了整个产业的运营协同, 形成了连接企业内部生产单元和企业外部合作伙伴的数字生产网络, 为客户提供了更优的体验。

综上, 发展数据要素市场是中国数字经济高质量发展的核心引擎。通过着力提升数据供给质量、破解数据流通障碍, 释放数据要素的乘数效应, 加快培育新技术、新模式、新业态, 数据要素将为经济社会各领域高质量发展提供有力支撑。

(三) 创新是数字经济的驱动力

创新是数字经济的重要驱动力, 其核心特征是平台借助于大数据、云计算、人工智能和移动通信等数字技术, 推动模式创新, 颠覆传统产业。模式创新通过重新定义产品或服务的生产、交付和获取方式, 打破了传统行业的界限, 不仅催生了电子商务、共享经济和订阅服务等商业模式创新, 还极大地重塑了市场的时空边界和组织结构形态, 创造了新的市场机会和价值。

基于大数据和数字技术的创新模式正从聚合走向分散^[20]。聚合的创新模式体现在集中的地理和资源分布上, 更加突出精英和专家的决策作用。然而, 在面对数据量庞大、参与人员分散和行为不可控等因素时, 通过建立协同平台和沟通媒介降低链接成本, 分散合作模式甚至可以产生超越精英和专家的创新成果。分散合作模式的成功, 有赖于开放权限, 旨在让更多的人参与创新; 有赖于自我组织, 旨在让参与者基于兴趣和特长选择承担的任务, 而非强行指派; 有赖于去名利化的追求, 旨在让参与者基于热爱主动作出努力, 而非受追求功利的驱动被动接受; 有赖于贡献的可证实与可撤销, 旨在确保成果的质量达到一定标准; 有赖于由技术领袖决定技术标准和引领方向, 旨在维持分散的管理结构。

相较于海外同侪, 国内缺乏具有影响力的开源技术。在数字经济背景下, 优秀的开源项目可以吸引更多开发者加入, 凝聚成极具创新力的开发社群, 并对市场需求作出及时反馈。例如, 谷歌开源的 Kubernetes 帮助 Linux 系统以极低的成本提高运行性能和可靠性; Meta 发布的 React 以出众的性能、简单快捷的优势成为炙手可热的前后端开发引擎; 谷歌的 Android 已形成成熟的生态, 拥有数量庞大的应用程序, 也孵化出了小米 MIUI 和华为 EMUI 等定制系统。由于中国开源生态建设整体起步较晚, 因而在开源项目的培育和孵化、开源社区运营和开源人才培养等方面与国际先进水平尚存在一定差距。

近年来, 中国平台企业创新效果欠佳。以 GPT 为代表的大模型问世是人工智能发展历史的里程碑事件。2020 年 5 月, 美国人工智能研发公司 OpenAI 异军突起, 发布语言模型 GPT-3, 短时间内便席卷全球。市场规模优势虽然为中国数字经济平台企业的发展提供了巨大的需求, 但也导致这些企业仍选择投入大量资本用于短期流量竞争, 对中长期竞争所需要的技术投入与创新探索不足。例如, 中国行业排名前 15 位的平台企业研发规模与美国行业排名前 15 位的平台企业相比存在明显差距, 其基础与底层技术创新能力亟待加强。创新能力欠缺导致 2021 年 2 月以来中国头部平台企业的估值大幅下滑, 中美头部平台企业市值差距急剧扩大。

(四) 个性化服务是数字经济的抓手

中国数字经济的快速发展离不开商业模式的颠覆式创新和转型, 这使消费者可以在任何时间与空间内表达个性化需求和偏好。平台企业运用自身优越的数据分析与算法能力, 开发出大量差异化、个性化的产品或服务, 培育出社交电商、网约车和外卖等新平台类型, 极大地满足了消费者的个性化需求。

一个好的个性化服务需要解决传统经济学理论中生产什么和为谁生产这两大核心问题, 代表了数字经济商业的未来。例如, 在虚拟世界中, 元宇宙允许每个用户进行内容生产和场景编辑, 通过赋予用户前所未有的创造力和自由度, 为创新提供广阔的空间。与此同时, 平台能够深入理解用户的兴趣和行为模式, 进而提供定制化的产品或服务。Sun 等^[21]研究发现, 取消个性化推

荐会导致交易额大幅下降81%。可见，精准的个性化推荐不仅可以提高用户的满意度和忠诚度、为平台带来更高的转化率和利润，还能够促进厂商进行产品创新和质量提升^[22]。因此，个性化推荐在很多数字经济平台的商业模式中扮演了重要角色。

一方面，个性化服务的日渐普及反映了数字经济时代长尾需求的充分释放。长尾需求的特点是多样化、个性化，它代表了消费者对独特产品或服务的需求。在传统经济中，由于成本和效率等方面的约束，企业往往无法顾及这部分个性化、零散、少量的需求。虽然对单个独特商品或服务的需求所占的市场份额很小，但总体数量加起来也能形成一个庞大的市场。而借助于数字技术，企业可以通过关注和分析长尾需求，开发满足这些需求的产品或服务，从而开拓新的市场、吸引新的流量。另一方面，个性化服务虽然能帮助平台吸引流量和鼓励创新，但也可能产生大数据杀熟、用户上瘾、信息茧房和回声室效应等问题^[23]。过度个性化的服务可能会导致用户沉浸于单一的信息环境中，从而限制其视野和认知。此外，个性化推荐系统可能会不断强化用户的既有偏好，甚至产生依赖性，这些都是个性化服务需要警惕和解决的问题。为此，《中华人民共和国个人信息保护法》第二十四条明确规定：“不得对个人在交易价格等交易条件上实行不合理的差别待遇”“通过自动化决策方式向个人进行信息推送、商业营销，应当同时提供不针对其个人特征的选项，或者向个人提供便捷的拒绝方式”。

综上所述，中国的数字经济正处于从数字化到数智化的关键转折时期。在数字化时代，企业积累了大量的数据，但这些数据往往因为缺乏有效的处理和分析而被闲置。而数智化则能够以海量大数据为基础，结合人工智能及相关技术（如机器学习、深度学习、自然语言处理、计算机视觉等）实现智能化决策。本文认为，智能化决策将越来越贴近经济学中的理性决策（Rational Decision）。经济学模型往往假设决策者具备最优化（Optimization）、完美回忆（Perfect Recall）和理性预期（Rational Expectation）等能力。然而，现实世界和实验室研究发现，人类行为与这些理论预期之间经常存在偏差。随着智能化决策系统的广泛应用，经济活动的结果将更加符合基于理性假设的经济学模型的预期，使这些模型在实际应用中更加精准和有效。因此，数字经济的新商业模式虽然意味着传统经济学理论分析框架在某些领域不再适用，但也使传统经济学中理性人假设越来越符合现实。

人工智能技术在各个行业的实践与应用都体现为通过数智化赋能企业决策。在广告行业，谷歌AdWords和Facebook Ads通过人工智能算法实现了更高的点击率和转化率。亚马逊、沃尔玛和京东等电商平台为企业提供智能履约服务，帮助企业预测需求、优化库存管理和物流，降低经营成本。通过人工智能技术，数智化能够对数据进行深入挖掘和分析，从而深刻改变企业的决策方式，使决策更加精准、高效，并能够实时响应市场变化。

四、数据要素市场发展新问题

数据成为数字经济的核心生产要素有赖于其专业性、稳定性、合法可流通性，以及与算力、算法和应用场景的协同性。数据要素化不仅可以优化资源配置、提高生产效率，还能为经济发展提供驱动力。数据要素化离不开数据要素市场的发展，但数据要素市场发展仍面临许多挑战。

传统的生产要素主要包括土地、劳动力、资本和技术等。伴随着经济的发展、经济学理论的进步和数字经济的日益繁荣，人们通过经济活动产生的数据对生产和服务起到的作用也愈发重要。然而，数据要素化需要满足以下条件：首先，数据要素的产权需清晰界定，相关权责归属明确；其次，数据要素需要在地区间、企业间、部门间能够以“可用不可见”的方式自由流动，且不会因非合理制度因素的限制而产生数据孤岛问题；最后，数据要素需要在一个充满活力、健康有序的要​​素市场中发挥价值，市场价格需能充分反映其在生产过程中的边际贡献，并准确传递数据要素的供求情况和稀缺性信号，进而指引市场主体对要素进行合理配置。

发展数据要素市场的过程是将尚未完全由市场配置的数据要素转向由市场配置的动态过程, 其目的是形成以市场为根本的调配机制, 实现数据流动的价值或使数据在流动中产生价值^[24]。数据要素的新特性意味着数据要素市场发展需要解决如下新问题:

第一, 数据要素的价值和报酬难以匹配。数据要素具有非竞争性且可以无限复制、重复使用。这意味着一份数据可以被不同企业同时使用。而其他生产要素通常不具有这种特性。一方面, 非竞争性意味着数据相较于其他的生产要素能被更广泛使用, 从而创造更多价值。另一方面, 数据一经出售, 就很难防止买方进行分享或二次转售, 从而损害数据所有者的权益。

第二, 数据要素的生命周期复杂。数据要素是经济活动的副产品, 数据要素产生之后还要经历采集、储存、清洗、加工、分发、传输、使用和消除等整个数据生命周期。数据要素的副产品特性导致其不同于石油、煤炭等自然资源, 其是可以源源不断地产生的。正因如此, 数据要素市场必须对数据要素生命周期进行全覆盖。

第三, 数据要素的自主控制权显著弱于其他生产要素。对于劳动力或资本要素, 其使用往往都要得到所有者的授权。但数据非常容易在未授权的情况下被使用, 因而数据要素市场的健康发展离不开对数据的规范使用和对数据隐私安全的严格保护。

第四, 数据要素价值的高度异质性和不确定性阻碍了数据的流通。一份数据在某个人工智能模型中能产生的价值依赖于使用的具体模型、模型中的其他数据以及具体的应用场景等, 具有高度不确定性。对此, 应鼓励市场主体探索更灵活的数据交易模式, 尽可能消除买家对数据价值的疑虑, 促进数据流通业务的推广和落地。

五、数据要素的价值创造与分配

数据要素市场发展面临的主要问题可以总结为如何构建符合数据新特性的数据要素价值创造与分配机制。数据要素的价值创造与分配是密不可分的。其中, 数据要素的价值创造是指不同市场主体各司其职, 通过提供、开发、建模和个性化服务等手段深入挖掘数据要素价值。而数据要素分配作为一种激励制度, 可以最大程度地激发市场主体的活力、充分发挥和利用数据价值。只有更好地解决数据分配问题, 才能从根本上促进数据的价值创造。本文绘制了数据要素价值创造与分配拓扑图, 具体如图3所示。

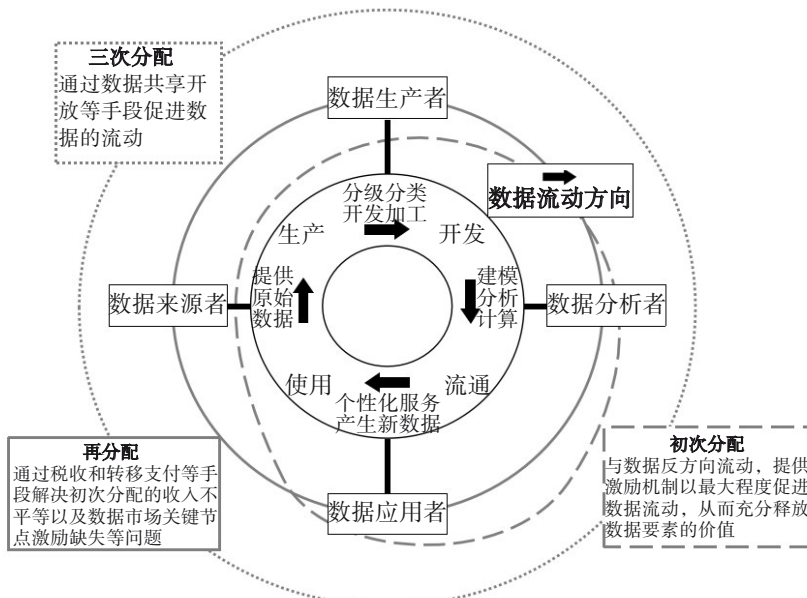


图3 数据要素价值创造与分配拓扑图

由图3可知,数据从来源者、生产者、分析者到应用者的流动,反映了数据的价值递增过程。数据价值创造包括生产、开发、流通和使用四个阶段,即来源者提供原始数据,由生产者开发与加工,经过分析者建模与计算,通过应用者提供个性化服务,再次产生新数据。与之相反的方向则是数据价值的初次分配过程,即市场通过提供合理的激励机制促进数据的产生与流动,以此释放数据要素价值的过程。

(一) 数据要素的价值创造

当前阻碍数据要素价值创造的主要因素之一是难以衡量数据要素的收益和贡献。当数据要素的价值以及各方的贡献无法被合理评估时,就会降低各方参与的积极性。因此,本文基于数据要素的经济学新理论框架对如何确定数据要素的收益和贡献提出可行的建议。

本文认为,数据要素收益的确定可以分为中心化和分散化两种方式。无论采用哪种方式,都需要创新交易机制,进而设计出符合数据要素特性的收益确定方式。例如,在一个数据联盟中,多个数据提供商提供不同数据集,每个数据提供商的收益确定就可以采取中心化的方式。在此情况下,一个简单的、直截了当的方式就是依据数据要素的贡献决定其收益。但这种方式可能会导致数据提供商的收益无法弥补其成本,从而降低参与积极性。Yu等^[25]认为,基于懊悔(Regret)函数动态调整数据要素收益的交易机制可以保证数据提供商有激励参与。

此外,还可以通过拍卖等分散化的方式决定数据要素收益。但由于数据具有非竞争性,以“出价最高者得”的拍卖机制来交易数据显然不是最有效的资源配置方式。此时,可以设计一种新的公开增价拍卖机制:数据提供商先依据数据成本设定目标总收益;在报价不断上升的过程中,依据使用的模型、模型中的其他数据以及应用场景等信息,每个数据竞价者可以计算数据要素贡献并与当前价格进行比较,当数据要素贡献低于当前价格时,数据竞价者退出拍卖,否则继续参与;在每个报价之下,都有一定数量的数据竞价者选择参与,数据提供商总收益即为价格乘以数据竞价者的数量,当数据提供商总收益达到了目标总收益时,拍卖停止;最终所有选择继续参与的数据竞价者都能够使用数据,所需支付的价格为拍卖停止时的价格。

传统研究主要采用“Leave-One-Out Test”(LOO方法)衡量数据要素的贡献。其核心思想是比较使用全部数据集训练模型得到的预测精度和使用减去某些数据点的数据集训练模型得到的预测精度。两个精度之差即代表了减去的数据点的贡献。在大数据时代,LOO方法可能会产生较大偏差,因为相比于一个很大的数据集,每个数据点的边际贡献近乎为0。最新理论认为,用经济学上的Shapley值衡量数据要素贡献^[26-27]。Shapley值被认为是一种公平计算贡献值的方法。该方法对于数据集中所有包含某些数据点的子集,均采用LOO方法计算这些数据在子集中的边际贡献,然后计算所有边际贡献的平均值,即得到了这些数据的Shapley值。在实际应用中,当使用多个不同数据集联合训练模型时,如果需要计算一个大数据集中每个数据点的Shapley值将会非常复杂,但每个数据集的Shapley值非常容易确定。如果假定每个数据集中的每个数据点的数据质量相同,则用该数据集的Shapley值除以数据点的数量,可以得到每个数据点的Shapley值。Shapley值作为一种公平准确的评价数据贡献的方法,有助于提高数据要素供给的数量和质量,从而促使数据要素创造更大价值。

(二) 数据要素的分配方式

在数据要素市场的发展过程中,分配应处于核心地位。探讨数据要素参与分配的方式需要区分数字经济传统分配模式和允许数据要素流通交易下的新模式。

在传统模式下,数据要素参与分配的方式主要来自于数字经济平台采取的交叉补贴定价策略。数据垄断是平台维持垄断地位的重要手段,平台可以用零价格甚至通过发红包和优惠券等方式招揽用户来获取数据。2014年诺贝尔经济学奖得主Tirole从博弈论和多边市场角度作出解释:在双边市场中,不同于以往中介低买高卖的盈利模式,平台可以采用更加灵活的诸如非对称定价

和交叉补贴等定价策略, 通过调整收费结构, 对一方免费以吸引用户, 对另一方收费以获取利润, 最终实现规模和利润的增加。因此, 社交、搜索、新闻和视频等平台普遍采用在线广告的方式赚取利润, 即“羊毛出在狗身上猪来买单”。在这种定价模式下, 个人数据所有者可以享受免费服务甚至获得红包和优惠券获得一次分配收益。但同时也应注意到, 因为信息不对称和隐私保护力度不够, 个人数据所有者获取的收益非常低。在允许数据要素流通交易的情形下, 数据要素将通过以下方式参与分配:

1. 初次分配

初次分配具体方式的选择取决于如何对数据要素进行合理估值和定价, 如何平衡数据来源者、数据处理者和数据应用者等多方利益, 以及如何实施等问题。

当数据要素可以进行流通交易时, 在传统的交叉补贴换取数据之外, 应当更多地发挥数据要素市场的分配作用。初次分配的目标应在保护个人数据安全使用的基础上, 深入挖掘数据价值, 厘清数据要素在生产活动中发挥作用的链条。初次分配过程中最核心也最困难的环节是分配主体的选择, 即数据要素的所有权或使用权的界定问题。虽然隐私计算技术的发展使得我们可以绕开数据所有权的问题, 单独讨论数据搜集者的使用权, 但数据的个人提供者理应作为分配主体参与数据要素的收益分配。一方面, 需要有相应的数据确权政策来保证数据个人提供者的合法权益。另一方面, 由于数据供需双方对个人隐私的价值评估存在较大信息不对称, 从而使得个人数据交易价格的合理制定存在极大困难, 因而需要创新性地设计相应的定价机制。

数据要素由于具有非竞争性、规模报酬递增等特性, 可以为提高生产效率、增加经济产出作出重要贡献。然而学者们发现, 相较于小企业和劳动者, 在生产中使用数据要素可能更加有利于大企业。Aghion 等^[28]研究发现, 20世纪90年代, 美国信息技术创新使科技企业快速扩张, 导致小企业活力及普通劳动力收入比重下降。Farboodi 等^[29]研究发现, 数据要素帮助企业提高产出后, 企业可以增加投资来扩大规模, 进而获取并积累更多数据资产。在这个过程中, 企业规模会出现两极分化。除非新进企业对数据要素使用效率更高并能够在早期获得融资, 才能通过不断试错积累数据, 进而逐渐超越现存大型企业。此外, Jones 和 Tonetti^[30]研究发现, 具有数据优势的企业会减少数据流通、阻碍市场进入和损害个人隐私, 进而降低整体收入和消费水平。

这些研究对于数据要素初次分配的启示在于: 在企业层面, 需要打通中小企业和初创企业获取数据要素的渠道, 同时提高中小数据要素使用效率, 以及为初创企业提供融资; 在消费者层面, 通过数据信托机制或设立个人及企业数据账户, 赋予微观主体行为数据的产权; 在数据需求方和供给方层面, 探索以数据开发或增值服务置换数据提供方的股权或特定数据权益的“增值入股”“数据入股”等方式, 帮助更多中小企业及个人在数据增值中受益。

2. 再分配

由于数据要素的初次分配主要集中在少数大型企业, 因而政府针对数据要素进行的再分配就对市场效率和维护公平意义重大。目前数据要素的再分配手段主要为, 与数据要素或数字经济挂钩的直接税(企业利润)和间接税(消费)。

自2018年欧盟提出数字税想法以来, 根据用户数、数字服务合同和数字经济活动收入等对一定规模的大型互联网企业征收数字税已经成为全球各国的普遍共识。2021年OECD发布的《关于应对经济数字化税收挑战“双支柱”方案的声明》使大型跨国企业将在其主要消费市场承担更多纳税义务。欧盟国家征收数字税的经验虽然对中国有一定的借鉴意义, 但也不能照抄照搬。欧盟国家征收数字税的考量主要关注区域间的征税权分配, 以及数字经济收益在跨国互联网巨头与中小消费国之间的分配, 因而受国际贸易规则和数据主权博弈的影响较大。其设计来自于对现有国际利润税体系进行改革, 并提出按收入的剩余利润分配(RPAI)和基于目的地的现金流量税(DBCFT)机制。相比之下, 中国的数字税设计则需要考虑在全国统一大市场下, 在大企业与小

小企业之间、发达地区与欠发达地区之间、资本回报与劳动报酬之间进行合理的分配。

关于在微观层面调节企业和消费者再分配的税收设计,需要考虑扭曲企业决策、导致利润和无形资产转移^[31]、打击企业对数据要素开发投入的积极性、激化税收竞争或报复等潜在影响。Cui^[32]研究发现,数字税可能优于其他类型消费税。增值税和基于目的地的现金流量税等可能缺乏对分配中公平问题的考虑。Schoen^[33]提出,数字税可借鉴企业所得税对特定“数字投资”的投资回报收税。Olbert和Spengel^[34]提出,数据要素作为数字经济中的新型价值驱动资产,应采用转让定价方法调节分配。另外,政府还可扩大对数据共享开放、数据基础设施建设等领域的财政投入,促进数据要素再分配。加大对数据共享开放的财政支持,以充分发挥政府数据的公共品作用,有效补充企业自有数据,从而优化企业的商业决策。

3.第三次分配

数据要素的非竞争性使其可以通过数据共享实现第三次分配。可通过鼓励行业协会、企业联盟和科研院所等社会组织与非营利性机构,开展数据共享开放、技术研发、权益保护和补偿等工作。支持搭建数据开源平台,鼓励公众、企业和社会各界创建与维护数据开源项目,以降低重复投入、促进广泛使用并持续推动创新。政府部门或行业协会可以通过设立奖惩机制推动企业进行数据共享。但是在某些情况下,强制性的制度会降低企业共享数据的积极性。数据的共享开放虽然产生了巨大的价值,但在共享开放的过程中,由于部门制度或本位主义的限制,会造成数据壁垒,进而形成数据孤岛现象,带来多方面的福利损失。

(三) 数据要素的收益分配主体

数据要素分配问题的复杂性在于需要考虑多方主体的利益。健全数据要素分配制度需要进一步厘清各个市场分配主体所扮演的角色,以及在不同主体之间分配多少、如何分配,进而充分激发市场主体活力。

1.企业主体

企业主体包括自有数据持有者、专业化数据商和数据增值服务提供商等。不同企业主体对数据分配的诉求存在巨大差异。其一,自身已持有一定规模数据的企业(例如,大型互联网平台企业等)既是数据持有者又是数据使用者,同时还是数字经济传统分配模式下的既得利益者,其参与数据共享或交易可能对其市场地位产生不利影响或导致利润减少,但若强制其参与数据分配,可能导致其减少对数据资源开发的投入^[30, 35]。社会层面的分析则表明,数据分配能促进更多企业适应消费者需求并使个人数据所有者获利,从而改善整个社会的分配格局^[36]。为了进一步找出数据持有者愿意参与数据分配的条件,有研究发现,当数据持有者是风险厌恶的且其他数据交易买方是风险中性时,持有者倾向于销售数据产品以平衡风险分摊和竞争加强的影响,否则数据持有者将选择销售数据和自留数据参与后续竞争的混合策略^[37-38]。其二,相比于自有数据持有者,专业化数据商主要对不同来源的原始数据或脱敏数据进行采集、存储、汇聚和加工,形成更加多元和综合的数据供给,并能有效促进消费者与市场主体间的数据流通^[39]。当数据需求存在网络外部性,使用者越多、价值越大时^[40-41],数据持有者可以通过提供差异化产品销售高质量数据,以实现价值最大化。独立的数据商并不使用数据,避免了与数据持有者形成竞争,因而数据持有者愿意通过向数据商支付渠道费用来增加销量。当企业利用数据进行创新存在不确定性时,独立数据商的存在有利于数据买方中和风险,通过补偿创新失败的损失增加企业创新的激励^[4]。专业化数据商更有激励通过数据要素市场参与分配,创造出更多惠及普通劳动者的新产业。

此外,市场中还存在提供模型化和人工智能化决策等增值服务的企业。这些企业普遍面临市场对智能化决策存有疑虑、业务推广难和落地难等困境。对于这些企业而言,应允许其探索更灵活的分配方式,如“增值入股”(持有一部分被服务企业的股权,从而享受被服务企业数据增值的收益)或“数据入股”(允许以数据增值服务置换被服务企业的特定数据权益)等,以打消市

场对使用数据的疑虑。

2. 第三方机构

数据要素市场中还需要大量第三方机构提供数据评估（质量、资产价值、安全）、公证、审计、培训和认证等服务。其中，数据资产价值评估是最核心的工作。

传统资产价值评估的成本法、收益法和市场法在应用于数据资产估值时各有其适用性和局限性^[39]。成本法操作简单但测算结果可能低于数据的实际使用价值^[42]。收益法需要对数据资产进行评估，但仍面临如何折现的问题。市场法主要基于历史交易价格，但无法在市场起步阶段作为可行的方法。市场实践中现有的数据收费模式包括：以一部分免费数据换取成交量和市场份额^[43]、提供一定价值的其他产品或服务获取消费者个人数据的授权^[44]、根据用量计费或收取订阅和租赁费的混合费用^[45-46]等。如何针对数据要素“有效期短、可以无限共享、集合使用价值更高”等特性设计最有效的价值评估方法是未来需要重点解决的问题。

3. 政府部门

通过政策法规强制企业参与数据要素市场不仅存在额外的行政成本，还可能引发企业降低数据质量等策略行为，从而减少社会福利^[47-48]。因此，政府在数据要素市场中应该扮演引导而非主导角色。政府主要通过以下两种方式参与数据要素市场：

第一种方式是政府以公共数据作为抓手直接参加市场流通。公共数据是指以政府和被政府授权的负有公共事务管理职能的非政府组织在处理公共事务过程中获取的数据。政府的公共数据可以通过共享开放承担更多的公益职能。然而，政府数据的共享开放仍有许多亟待解决的问题。如最优共享开放模式是什么？应共享开放哪种数据（如基础数据）？面向哪些用途场景（如抗疫救灾）和数据需求方（如公益性机构和小微企业）共享开放？是否收费及如何收费（如一定周期内一定用量免费、大企业收费小微企业不收费、深度加工融合后通过“市场化机制”有偿提供但比市场费用低等）？通过借鉴信息经济学中的公共信息理论，可以对上述问题进行回答。现有文献中提出了许多公共信息披露的优劣性。例如，当存在协调问题时，越精确地披露公共信息就越能增加社会福利^[49]。但公共信息披露会有挤出效应^[50-51]，公共信息披露过多会导致个体获取信息的动力不足；公共信息披露在有些情形下还会导致个体交易机会的丧失^[52]。因此，最优的公共数据共享开放需要综合考虑上述优劣性。另外，通过探索授权运营、数据出让和数据采购等创新制度，有利于促进政府数据的利用与开发，实现国有资产保值增值，促进社会数据供给与政—企数据融合利用等^[53-54]。综上所述，政府的公共数据更多地可以通过其公益性质参与第三次分配，但也可以探索如何通过提供增值服务获取一定的初次分配收益。

第二种方式是政府部门可以为数据要素市场发展创造政策环境，更好地激发微观市场主体的活力。特别是对于数据商和第三方机构，政府可以采用补贴和税收优惠等方式激励更多企业参与数据要素市场活动。但在具体政策设计上，应警惕强制参与可能导致的额外行政成本和策略行为，以及由于信息不对称导致的骗取政策优惠等问题^[55]。未来需要针对数据要素市场的具体场景，设计满足激励相容约束的政府最优优惠政策组合，借助财政、税收、融资、上市、信托、保险和信贷等政策工具激活市场主体参与的积极性。

本文认为，数据要素分配的方式应是在按劳分配和按生产要素分配并存的前提下，建立强调效率属性的初次分配制度和突出公平属性的再分配制度。在初次分配方面，应首先通过合理手段确定数据要素的贡献和收益，使数据要素参与分配的额度与数据要素在生产价值创造过程中的贡献率相符合，并在数据来源者、生产者、分析者和应用者等不同市场主体之间科学决策分配给谁、分配多少以及如何分配等问题。在再分配方面，应探索数字税或利润税改革，实现数据红利在大企业与中小企业之间、发达地区与欠发达地区之间的再分配；同时也可探索以数据开发或增值服务置换数据提供方的股权或特定数据权益的“增值入股”“数据入股”等方式，帮助更多中

小企业及个人在数据增值中受益。在第三次分配方面,既应推动政府公共数据承担更多公益职责,也可鼓励行业协会、企业联盟和科研院所等社会组织与非营利性机构,开展数据共享开放、技术研发、权益保护和补偿等工作。支持搭建数据开源平台,鼓励公众、企业和社会各界创建与维护数据开源项目,减少重复投入,促进数据要素广泛使用,并持续推动创新。

六、结 论

本文立足于博弈论与信息经济学理论构建了一个数据要素的经济学新理论框架,该框架以数据为核心生产要素,涵盖了从流量、数据、创新到个性化服务的数字经济循环。与此同时,数据要素化和数据要素市场的发展仍面临如何构建合理的数据要素价值创造和分配机制等问题。基于本文提出的数据要素的经济学新理论框架,本文尝试解答了关于数据要素收益和贡献的确定、数据要素参与分配的方式与主体等数据要素市场发展中的关键问题。

当前数据要素的经济学新理论框架还有如下需要进一步研究的问题:其一,现有的博弈论与信息经济学的理论框架一般都假定博弈参与者是完全理性的,但现实中人们往往存在诸如偏好不一致等行为偏误问题。现有理论框架还无法很好地分析存在行为偏误下的博弈问题。其二,上述理论框架是基于数据要素权属界定清晰的前提下对价值创造和分配进行分析,但现实中数据权属界定不清已成为数据要素化主要的制度障碍之一。当市场存在交易成本时,数据归属于平台、消费者或政府的不同初始配置将影响数据资源的最终配置和社会福利水平。应结合不同场景、不同参与者的博弈互动关系,完善数据界权规则,审慎界定个人、企业和国家的数据权责。其三,数据要素的经济学新理论框架尚未回答有关数据分级分类和价值评估等问题,未能基于数据要素的经济属性提出完整、明确的数据分类体系标准,并且仍需参照会计学中对传统无形资产估值方法,结合数据要素特性,探索出有效的价值评估方法。

着眼于现实,从数据要素市场破局的纲领性文件《“数据要素×”三年行动计划(2024—2026年)》来看,数据要素市场将迎来爆发式增长,而且地方政府发展数据要素市场的热情高涨。然而,数据市场作为一种崭新的市场形态,并没有任何成熟理论来指导其发展,尚处于竞相探索阶段。同时,从地方实践中也可以看到,数据面临着要素化、市场化和价值化的多种挑战。数据要素化需要解决数据确权、价值挖掘和形成稳定需求和供给等问题;数据市场化需要建立完善的数据市场体系、公平高效的收益分配机制并形成合理的交易价格;数据价值化需要解决数据资产入表评估问题,形成一套行之有效的数据资产估值体系。在许多现实场景中,这些挑战相互交织。可见,数据要素市场的发展在鼓励地方先行先试的同时,也需要在发展中规范、在规范中发展,有很多理论问题和实践问题亟需学术界的深入研究。理论方面的问题包括如何构建纳入数据要素的新生产函数、如何理解数据要素与其他生产要素的协同联动机制及其对全要素生产率的贡献、如何构建数据要素生产、流通、分配等基础性经济理论等;实践方面的问题包括如何开展数据要素典型应用场景的案例总结与推广、如何设计数据要素使用和流通效率的评估方法、如何开展数据资源盘点和数据资产入表等。

参考文献:

- [1] DE CLIPPEL G, ELIAZ K, ROZEN K. Competing for consumer inattention[J]. Journal of political economy, 2014, 122(6): 1203-1234.
- [2] GABAIX X. Behavioral inattention [C]//BERNHEIM D, DELLAVIGNA S, LAIBSON D. Handbook of behavioral economics. Amsterdam: Elsevier, 2019: 261-343.
- [3] MOODY D, WALSH P. Measuring the value of information: an asset valuation approach[R]. Proceedings of the 7th European Conference on Information Systems, 1999.

- [4] AKCIGIT U, LIU Q. The role of information in innovation and competition [J]. *Journal of the european economic association*, 2016, 14(4):828-870.
- [5] 李三希, 曹志刚, 崔志伟, 等. 数字经济的博弈论基础性科学问题 [J]. *中国科学基金*, 2021, 35(5):782-800.
- [6] KAMEPALLI S K, RAJAN R, ZINGALES L. Kill zone [R]. NBER Working Paper No. 27146, 2020.
- [7] HOLMSTRÖM M, PADILLA J, STITZING R, et al. Killer acquisitions? The debate on merger control for digital markets [J]. *Yearbook of the finnish competition law association*, 2018(17): 35-54.
- [8] GAUTIER A, LAMESCH J. Mergers in the digital economy [J]. *Information economics and policy*, 2021, 54(C): 100890.
- [9] DE NIJS R. Behavior-based price discrimination and customer information sharing [J]. *International journal of industrial organization*, 2017, 50(C): 319-334.
- [10] FUDENBERG D, TIROLE J. Customer poaching and brand switching [J]. *The RAND journal of economics*, 2000, 31(4): 634-657.
- [11] CHOE C, MATSUSHIMA N, TREMBLAY M J. Behavior-based personalized pricing: when firms can share customer information [J]. *International journal of industrial organization*, 2022, 82(C): 102846.
- [12] SAMUELSON P A, NORDHAUS W D. *Economics* [M]. New York: McGraw-Hill Irwin, 2010:26-30.
- [13] 国家互联网信息办公室. 第45次《中国互联网络发展状况统计报告》[EB/OL]. (2020-04-28)[2024-03-01]. https://www.cac.gov.cn/2020-04/27/c_1589535470378587.htm.
- [14] 中国信息通信研究院政策与经济研究所. 平台经济与竞争政策观察(2020年)[EB/OL]. (2020-06-30)[2024-03-01]. https://www.199it.com/archives/1059581.html#google_vignette.
- [15] 唐维红, 唐胜宏, 刘志华. 中国移动互联网发展报告(2022) [M]. 北京: 社会科学文献出版社, 2022:2-9.
- [16] 月狐数据. 2022年Q1移动互联网行业数据研究报告 [EB/OL]. (2022-04-28)[2024-03-01]. <https://www.moonfox.cn/insight/report/986>.
- [17] 麦肯锡, 中国连锁经营协会. 2022年中国零售数字化白皮书 [R]. 2022中国国际零售创新大会, 2022.
- [18] NORDHAUS W D. Are we approaching an economic singularity? Information technology and the future of economic growth [J]. *American economic journal: macroeconomics*, 2021, 13(1): 299-332.
- [19] FARBOODI M, VELDKAMP L. A model of the data economy [R]. NBER Working Paper No.28427, 2022.
- [20] MCAFEE A, BRYNJOLFSSON E. *Machine, platform, crowd: harnessing our digital future* [M]. New York: W.W. Norton and Company, 2017.
- [21] SUN T, YUAN Z, LI C, et al. The value of personal data in internet commerce: a high-stakes field experiment on data regulation policy [J]. *Management Science*, 2024, 70(4): 2645-2660.
- [22] AGUIAR L, WALDFOGEL J. Quality predictability and the welfare benefits from new products: evidence from the digitization of recorded music [J]. *Journal of political economy*, 2018, 126(2): 492-524.
- [23] ACEMOGLU D. Harms of AI [C]//BULLOCK J B, CHEN Y C, HIMMELREICH J, et al. *The oxford handbook of ai governance*. Oxford: Oxford University Press, 2022:660-706.
- [24] 翁翕. 加快推进数据要素市场化建设 充分发挥数据要素作用 [J]. *中国经贸导刊*, 2022(3):31-32.
- [25] YU H, LIU Z, LIU Y, et al. A sustainable incentive scheme for federated learning [J]. *IEEE intelligent systems*, 2020, 35(4): 58-69.
- [26] GHORBANI A, ZOU J. Data shapley: equitable valuation of data for machine learning [R]. *Proceedings of the 36th International Conference on Machine Learning*, 2019.
- [27] JIA R, DAO D, WANG B, et al. Towards efficient data valuation based on the shapley value [R]. *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, 2019.
- [28] AGHION P, ANTONIN C, BUNEL S. Artificial intelligence, growth and employment: the role of policy [J]. *Economics and statistics*, 2019(510-511-512): 149-164.
- [29] FARBOODI M, MIHET R, PHILIPPON T, et al. Big data and firm dynamics [J]. *AEA papers and proceedings*, 2019(109): 38-42.
- [30] JONES C I, TONETTI C. Nonrivalry and the economics of data [J]. *The American economic review*, 2020, 110

- (9):2819-2858.
- [31] DISCHINGER M, RIEDEL N. Corporate taxes and the location of intangible assets within multinational firms[J]. *Journal of public economics*, 2011, 95(7): 691-707.
- [32] CUI W. The digital services tax: a conceptual defense[J]. *Tax law review*, 2019, 73(1): 69-112.
- [33] SCHOEN W. One answer to why and how to tax the digitalized economy[J]. *Intertax*, 2019, 47: 1003-1022.
- [34] OLBERT M, SPENGLER C. Taxation in the digital economy: recent policy developments and the question of value creation[R]. ZEW Discussion Papers No.19-010, 2019.
- [35] CURRARINI S, FERI F. Information sharing in oligopoly[C]//CORCHÓN L C, MARINI M A. *Handbook of game theory and industrial organization*. Cheltenham and Northampton: Edward Elgar, 2018:520-535.
- [36] EASLEY D A, HUANG S, YANG L, et al. The economics of data[R]. Social Science Research Network Working Paper No.3252870, 2018.
- [37] ADMATI A R, PFLEIDERER P. Selling and trading on information in financial markets[J]. *American economic review*, 1988, 78(2): 96-103.
- [38] GRINBLATT M S, ROSS S A. Market power in a securities market with endogenous information[J]. *Quarterly journal of economics*, 1985, 100(4): 1143-1167.
- [39] 熊巧琴,汤珂.数据要素的界权、交易和定价研究进展[J].*经济学动态*,2021(2): 143-158.
- [40] GLAZER R. Measuring the value of information: the information intensive organization[J]. *IBM systems journal*, 1993, 32(1): 99-110.
- [41] TOMAK K, KESKIN T. Exploring the trade-off between immediate gratification and delayed network externalities in the consumption of information goods[J]. *European journal of operational research*, 2008, 187(3): 887-902.
- [42] GHOSH A, RORH A. Selling privacy at auction[J]. *Games and economic behavior*, 2015(91): 334-346.
- [43] LERNER J, PATHAK P A, TITOLE J. The dynamics of open-source contributors[J]. *American economic review*, 2006, 96(2): 114-118.
- [44] PEI J. A survey on data pricing: from economics to data science[R]. ArXiv Working Paper, No.2009.04462, 2020.
- [45] WU S, BANKER R. Best pricing strategy for information services[J]. *Journal of the association of information systems*, 2010, 11(6): 339-366.
- [46] BAILEY W, LI H, MAO C X, et al. Regulation fair disclosure and earnings information: market, analyst, and corporate responses[J].*The journal of finance*, 2003, 58(6): 2487-2514.
- [47] BUSHEE B J, MATSUMOTO D A, MILLER G S. Managerial and investor responses to disclosure regulation [J]. *Accounting review*, 2004, 79(3): 617-643.
- [48] COHEN L, LOU D, MALLOY C. Playing favorites: how firms prevent the revelation of bad news [J]. NBER Working Paper No.w19429, 2014.
- [49] ANGELETOS G, PAVAN A. Efficient use of information and social value of information[J]. *Econometrica*, 2007, 75(4): 1103-1142.
- [50] GOLDSTEIN I, YANG L. Information disclosure in financial markets[J]. *Annual review of financial economics*, 2017, 9(1): 101-125.
- [51] GOLDSTEIN I, YANG L. Commodity financialization and information transmission [J]. *The journal of finance*, 2022, 77(5): 2613-2667.
- [52] KURLAT P, VELDKAMP L. Should we regulate financial information?[J]. *Journal of economic theory*, 2015, 158(B): 697-720.
- [53] 张鹏,蒋余浩.政务数据资产化管理的基础理论研究:资产属性、数据权属及定价方法[J].*电子政务*,2020(9): 61-73.
- [54] 谢波峰,朱扬勇.数据财政框架和实现路径探索[J].*财政研究*,2020(7):14-23.
- [55] 李文健,翁翕,龚六堂.政府如何激励创新? ——基于委托—代理理论的研究[J].*经济学(季刊)*,2022,22(2): 365-384.

A New Theoretical Framework for the Economics of Data

CAI Si-hang, WENG Xi

(Guanghua School of Management, Peking University, Beijing 100871, China)

Summary: Information technology, represented by big data and artificial intelligence, is advancing rapidly, bringing about profound business model innovations and leading to an explosive growth and massive aggregation of data. Data is increasingly becoming a crucial strategic resource and a novel factor of production, and the development of the data market is becoming the core driving force for the high-quality development of China's digital economy. However, regarding data as a factor of production and circulating data in a market are seminal concepts and there are no existing theories to guide their development. Although the new business models in the digital economy suggest that conventional theoretical frameworks in economics are no longer applicable in certain areas, they also make the rationality assumption in traditional economics increasingly align with reality. In practice, the development of China's data market still faces various challenges.

This paper, based on a new theoretical framework for the economics of data, focuses on the new elements of the digital economy such as platforms, traffic, and data, and studies the principles and key questions in the development of the data market. This framework is mainly based on game theory and information economics, emphasizing that the data is different from other traditional production factors because it is non-rival and it is a by-product. The increasing returns to scale and network effects mean that the data can achieve greater value creation than the sum of its parts, and improve the total factor productivity of other factors through amplification, superposition, and multiplication. Currently, the development of the data market still needs to address issues such as how to measure the value and contributions of data reasonably in order to provide sufficient incentives for participation. Additionally, it is necessary to design and implement primary, secondary, and tertiary distribution policies that achieve more equitable outcomes. Encouraging various market entities to explore and improve data trading mechanisms and pricing schemes while redistributing benefits from data through optimal tax system designs and data sharing initiatives can facilitate the full release of the value of data and improvement in social welfare.

Compared with existing literature, the contribution of this paper lies in proposing a new theoretical framework that combines the game theory and information economics. It analyzes traffic, data, innovation, and personalized services as an organic whole, and summarizes the positive feedback loop embedded in the digital economy with data as its core production factor. By incorporating the new characteristics of data, this paper constructs a topological graph to illustrate an allocation and distribution system for data. In particular, the creation of value from data proceeds in a direction opposite to the flow of data production, and larger firms tend to have a greater share in the initial distribution of the data economy. Based on this topological illustration, we propose practical suggestions for measuring the value and contributions of data using mechanisms such as auctions and the Shapley value, and answers questions about the implementation and entities associated with the distribution of benefits from data. This provides not only academic support but also theoretical basis for relevant institutions and policies in the development of the data market.

Key words: data; data market; information economics; value creation and distribution

(责任编辑: 徐雅雯)

[DOI]10.19654/j.cnki.cjwtyj.2024.05.003

[引用格式]蔡思航, 翁翕. 一个数据要素的经济学新理论框架[J]. 财经问题研究, 2024(5): 33-48.