

数据经济学

第一章：数据经济学导论

陈希路

暨南大学经济学院

2026 年春

章节目录

- 1 第一节：什么是数据经济学？
- 2 第二节：数据经济学对经济学理论的拓展和贡献
- 3 第三节：全球视域下数据经济学的核心研究议题

- **背景**：随着信息技术飞速发展，数据成为重要的生产要素
- **本章核心的三个问题**：
 - ❶ **背景与概念**：数据经济学产生的背景、意义；数据和其他基本概念的辨析
 - ❷ **拓展与贡献**：数据经济学对传统经济学理论和方法的**拓展与贡献**
 - ❸ **议题与进展**：全球视域下数据经济学的核心研究议题及研究进展评述

第一节

什么是数据经济学？

什么是数据经济学？

数据经济学的直观定义

数据经济学是一门研究**数据在经济活动中的作用和影响**的学科。

- **涉及环节**：数据的收集、加工、分析、存储和利用
- **目标**：
 - 助力实现数据的合理使用
 - 最大化数据的潜在价值

数据技术与数据经济：产业规模

- 数据已成为新的生产要素和财富来源
- **全球规模**（2025 年数据）：
 - IDC 报告：全球大数据分析市场规模为 3947 亿美元
- **中国规模**（2024 年数据）：
 - **数字经济核心产业**增加值：140891 亿元，占 GDP 的比重为 10.5%
 - **数字技术应用业**增加值：61928 亿元（占比 44.0%）
 - **数字产品制造业**增加值：48145 亿元（占比 34.2%）
 - **数字要素驱动业**增加值：26519 亿元（占比 18.8%）
 - **数字产品服务业**增加值：4298 亿元（占比 3.1%）
- **结论**：数据产业已成为经济增长的重要推动力

数据技术与数字经济：政策背景

- 数据作为生产要素的定位：

- 基础性资源、战略性资源、重要生产力
- 构建以数据为关键要素的数字经济

- 关键政策：

- 党的十九届四中全会（2019 年 10 月）提出：
 - 健全劳动、资本、土地、知识、技术、管理、**数据**等生产要素由市场评价贡献、按贡献决定报酬的机制
- 数据二十条（2022 年）
- 国家数据局成立（2023 年）
- “数据要素×”三年行动计划（2024 年起）

案例解析：数据如何成为重要的生产要素

场景对比：传统出租车与智能网联出行

传统模式下，出租车司机依靠个人经验在街头寻找乘客，核心投入是资本即车辆和劳动即驾驶时间。

- 数据要素的引入：

- 平台汇聚海量用户的实时地理位置和出行需求数据
- 结合路况数据，通过算法实现供需的瞬间精准匹配

- 经济学意义：

- 数据极大降低了搜寻成本和空驶率
- 数据与算法直接参与价值创造，体现了按贡献决定报酬的机制

数据技术与数据经济：研究的发展历程

数据经济的定义

数据经济是指以数据为基础的经济体系，包括数据的生产、加工、流通和使用等环节。

- Carl Shapiro & Hal R. Varian: 《信息规则：网络经济的策略指导》
 - 数据重要性日益凸显
 - 数据分析和挖掘是企业竞争的核心能力
- Mayer-Schönberger & Cukier: 《大数据时代：生活、工作与大变革》
 - 一场技术革命，改变生活、工作与大变革
 - 需关注隐私安全、可解释性和透明度

数据技术与数字经济：数据经济学的确立

- 数字经济正在成为经济学研究重要的新领域
- 术语提出：Varian（2014）首次提出“数据经济学”术语
- 数据的双重属性（《经济学人》，2020）：
 - ① 像石油：一种自然资源，最有价值，可以拥有和交易
 - ② 像公共产品：应该尽可能广泛使用，以最大化创造财富
- 挑战：必须建立新的机构来反映这种紧张关系（类似知识产权）

案例解析：石油属性与公共产品属性的碰撞

石油属性：医疗健康数据

制药企业投入巨资收集和分析数百万份非结构化病历数据，用于训练疾病预测模型和研发新靶点药物。此时，数据具有极高的商业**排他性**和增值潜力。

公共产品属性：气象与交通数据

政府部门将宏观气象预警数据和城市公共交通实时数据向全社会开放。无数中小企业基于公开接口开发便民服务软件。数据被不同人反复使用，价值并未损耗，体现了**非竞争性**。

数据技术与数据经济：带来的理论创新

- 数据对经济全方位、深层次的改变将引发一系列理论创新
- 数据革命不仅是生产力革命，也将带来生产关系的变革
- 历史上的 5 次范式演变：
 - ① 机械革命 → 早期机械化工业范式
 - ② 蒸汽动力革命 → 蒸汽动力和铁路范式
 - ③ 电力革命 → 电气和重型工程范式
 - ④ 石油革命 → 福特制大规模生产范式
 - ⑤ 信息革命 → 数字经济范式
- 意义：研究数据带来的范式转变是经济学的重要价值所在

案例解析：从福特制到数字经济范式

● 福特制大规模生产范式：传统服装零售

- 基于历史销售经验，提前半年进行大批量标准化生产
- 然而，信息不对称会导致高昂的库存成本和滞销风险

● 数字经济范式：以跨境电商 SHEIN 为例的小单快反

- 算法实时抓取全球搜索热词和社交平台趋势数据
- 驱动柔性供应链进行极小批量生产，根据市场实时数据反馈决定是否追加生产
- 海量数据流动重塑了生产组织方式，使得供需匹配达到了前所未有的效率

数据的基本定义：概述

- **直观比喻：**数据更像是石油或阳光（无处不在、源源不断）
- **词源：**
 - 拉丁文 *datum* 的复数形式
 - 《新牛津美语词典》：由计算机操作、存储、传输的数量、字符或符号
 - 《新华词典》：计算机加工处理的对象
- **现状：**目前尚无统一共识，但可从三个维度理解

数据的基本定义：三个维度

维度一

数据是对现实世界的抽象，是现实世界的模型。

- **观点来源：** 国际数据管理协会、美国质量学会
- **定义：**
 - 对事实的表现（文本、数字、图像等）
 - 对真实世界的对象、事件和概念的抽象表示
- **分类：**

类型	示例
结构化数据	数字
非结构化数据	文本、图像、音频、视频

数据的基本定义：三个维度

维度二

数据本质上是一种表示方式，要遵循特定的规范和标准。

- **来源**：国际标准化组织
- 数据是对象的解释，同时需要被解释
- 需要遵循特定的**规范和标准**以确保数据的权威性、通用性

维度三

数据可支持分析、推理、计算和决策，真实、准确是对数据的基本要求。

- **来源**：17 世纪哲学家、《新牛津美语词典》
- 科学界：建立知识、检验假说
- 企业界：提供产品、控制风险
- 政府界：提供公共服务、制定政策

概念辨析一：数据 vs. 信息

数据与信息区别

数据是信息的**基本构成单元**，而信息是对数据进行**加工处理**后所得到的有意义的结果。

- **数据 (Data)**

- 未经加工、处理或组织的原始事实
- 本身没有意义，是无序的
- **例子：**一组数字、文本文件、原始图像

- **信息 (Information)**

- 有用的、有组织的数据
- 具有含义和价值
- **例子：**销售趋势图表、分析报告

概念辨析二：数据 vs. 想法

数据与想法的区别

想法和数据都是信息的类型，但想法是一个生产函数，数据是一个生产要素。

- **想法 (Idea) $\approx$ 生产函数：**
 - 是一组制造商品的**指令**
 - 例子：自动驾驶的**算法**（非线性模型的预测规则）
- **数据 (Data) $\approx$ 生产要素：**
 - 是用于产生想法或在生产中使用的**输入**
 - 例子：传感器（相机、雷达、GPS）采集的**记录**

概念辨析二：数据 vs. 想法

区分“数据”与“想法”的经济学意义，主要在于**排他性程度**的认定：

① 传输难度：

- 数据：技术上更容易传输（互联网发送）
- 想法：需要投入大量资源去学习

② 排他性与监控：

- 数据可以加密特别是大数据更容易被监控，制成高度排他性的资产

自动驾驶领域的现状

机器学习的“想法”（算法）往往是公开的，但输入的“驾驶数据”是保密的。这导致了数据的**有限排他性**。

第二节

数据经济学对经济学 理论的拓展和贡献

数据经济学对经济学理论的拓展和贡献

核心观点

数据的引入，拓展了传统经济学理论的研究范畴。

- **挑战：**
 - 数据的**非竞争性**（一个人使用不影响他人使用）
 - 数据的**有限排他性**（难以完全阻止他人获取）
 - 这些特征挑战了传统经济学的经典假设
- **契机：** 为建立和发展数字经济时代的经济学理论带来了重要机会

对增长理论的拓展

- 传统基准：Solow 模型 (1957)

$$Y = AF(K, L)$$

- 数据出现后的修正尝试：

- 数据是信息：用于减少预测误差，回报有界
- 数据是劳动力：科技公司的关键投入
- 数据资本：将数据资本从传统资本中分离

- 现有局限：

- 往往忽略了数据与劳动、资本在经济性质上的差异
- 忽略了数据通过制度、规则、决策、协调等发挥的促进作用

对增长理论的拓展

- 内生增长理论 (Romer, 1990)

$$Y = AF(K, H)$$

其中 H 代表知识资本 (非减少性资源)

- 数据驱动经济的特殊性: 它是普遍内生增长模型的一个特例

- 具有以下结构性特征:

- ① 普遍的信息不对称
- ② 赢家通吃导致“超级巨星”公司增加
- ③ 出现未被传统会计系统捕捉的新形式贸易
- ④ 信息基础设施漏洞导致的系统性风险

- 机器智能的影响:

- 机器知识资本既是人力资本的补充, 也是替代品
- 可能导致人力资本价值的贬值

案例解析：机器知识资本对人力资本的补充与替代

场景：自动驾驶技术（Autonomous Driving）

在传统模型中，驾驶技能属于司机的人力资本，需要耗费大量时间培训且难以快速复制。

- 数据的引入与内生增长：

- 自动驾驶公司收集海量人类驾驶数据（图像、雷达、操作记录）
- 算法通过学习这些数据，形成了机器知识资本

- 经济学影响：

- 替代效应：机器知识资本使得职业司机的门槛降低，传统驾驶技能（人力资本）面临贬值
- 赢家通吃：拥有最多数据和最强算法的公司（如 Tesla、百度 Apollo）将形成难以逾越的壁垒，催生“超级巨星”企业

案例解析：大语言模型与机器知识资本

场景：生成式人工智能（如 ChatGPT、Gemini、DeepSeek）

传统经济增长模型中，文本创作、代码编写和逻辑推理高度依赖知识工作者多年积累的人力资本。

- **数据作为核心生产要素的具象化：**

- AI 企业汇聚互联网级别的海量高质量语料库和代码库进行预训练
- 算法将人类的隐性知识提取并转化为可复制、可扩展的**机器知识资本**

- **经济学影响：**

- **替代效应：**基础文案、初级编程等常规认知劳动面临快速贬值
- **互补效应：**成为顶尖专家的“副驾驶（Copilot）”，拓宽生产力边界
- **规模报酬递增：**拥有独家优质数据的企业，其机器知识资本的积累速度远超人力资本，进一步加剧了“超级巨星”企业的赢家通吃效应

对信息不对称理论的拓展

- 传统假设 vs. 现实：

- 完全竞争市场假设 → 信息完全对称
- 现实市场 → 信息分散、信息壁垒（阿克洛夫、斯蒂格利茨）

- 大数据的角色转变：

- 数据已从单纯的假设条件，变为新的生产要素
- 参与价值创造
- “信息不对称”形式演变为“数据不对称”

对信息不对称理论的拓展

1. 减少不对称（积极面）

- **降低搜索成本**：ICT 革命降低了商业周期影响
- **领域**：金融市场、劳动力市场、零售市场受益明显

2. 引发新型不对称（消极面）

- **人机不对称**：对人脑是噪声，对机器是信息矿
 - 机器提取系统信息的能力是其价值的基础
- **企业间/国家间**：巨额资本投资门槛、数字鸿沟

信息不对称的影响

信息不对称既是数据经济增长的源泉，也是它的“原罪”（导致市场失灵）。

案例解析：数据不对称与“大数据杀熟”

场景：平台经济中的一级价格歧视

数字平台（如电商、在线旅游 OTA、网约车）对新老用户展示不同的价格

- **数据减少了传统的信息不对称：**
 - 消费者可以瞬间比价，**搜索成本极低**（找酒店、买机票变得透明）。
- **创造了新型的人机不对称：**
 - 平台掌握了用户的浏览记录、消费能力、甚至是手机电量等海量数据
 - 消费者对平台的定价算法一无所知
- **原罪与失灵：**平台利用这种绝对的数据优势剥夺了消费者剩余，这正是数字经济可能导致市场失灵的经典表现

对交易成本理论的变革

- **传统交易成本**：由于信息不对称、合约不完备等，企业需承担搜索、谈判、监督、合规成本
- **数字技术的冲击 (Goldfarb & Tucker, 2019)**：降低了以下五种成本：
 - ① 搜索成本
 - ② 复制成本
 - ③ 运输成本
 - ④ 跟踪成本
 - ⑤ 验证成本

对交易成本理论的变革

具体表现为以下四个方面：

- ① **增加透明度：** 互联网快速传递信息，降低不对称程度
- ② **提高速度与效率：** 自动优化交易过程，减少人工介入，提高准确性
- ③ **新型交易模式与渠道：**
 - 模式：电子商务、虚拟货币、区块链（更安全透明）
 - 渠道：在线平台、APP
- ④ **企业边界扩大：**
 - 生产过程模块化、平台化
 - 涌现众包、外包、共享经济、零工经济
 - 生产主体更加多元化、微粒化

案例解析：交易成本骤降与企业边界的消融

场景：外卖平台与“零工经济”

传统餐饮企业需要雇佣专职配送员，需承担底薪、社保以及管理等高昂的**内部组织成本**。

- **数据对交易成本的降低：**

- **跟踪与验证成本：**GPS 实时定位和算法派单，平台能够以极低成本实时监督几百万骑手的工作状态

- **企业边界的革命：**

- 美团/饿了么通过平台化网络，将全社会的闲散劳动力纳入配送体系（众包/零工经济）
- **结论：**当依靠数据进行外部协调的市场交易成本低于企业内部管理成本时，企业的物理边界就会无限扩大，生产主体趋于微粒化

对经济学研究方法的拓展

- 科学研究的第四范式：

- 詹姆斯·格雷 (Jim Gray)：从实验 → 理论 → 计算模拟 → 数据探索
- 理论研究与应用实时融合

- 经济学的实证革命：

- 20 世纪 80 年代中期以来，实证研究论文比例攀升
- Angrist et al. (2017) 指出：顶刊中基于数据的实证论文从 <35% 上升至 55% 左右，理论性论文数量显著下降

对经济学研究方法的拓展

《大数据时代》(Mayer-Schönberger)

大数据时代的研究方法变革主要体现在三方面：

- ① **全体数据**：不再依赖样本，而是分析全体数据
- ② **接受混杂**：不再追求简单的精确性，乐于接受纷繁复杂的数据
- ③ **重相关，轻因果**：
 - 传统的因果关系难以验证
 - 更加注重容易捕捉的**相关关系**

案例解析：重相关轻因果——大数据预测

经典案例：Target 百货预测孕妇与谷歌流感趋势

大数据分析往往不深究“为什么”（因果），而是寻找“是什么规律”（相关）。

- **接受混杂与关注相关性：**
 - **Target 百货：**通过分析顾客购买无香型护肤品、大容量维生素等 25 种商品的消费记录，精准预测顾客是否怀孕，甚至推算出预产期
 - **谷歌流感趋势 (GFT)：**不依赖医学病理化验，而是通过分析各地区对“退烧药”、“咳嗽”等词汇的搜索频率，比疾控中心更早预测出流感的暴发趋势
 - 然而，GFT 后来因为算法衰减和忽视因果关系，导致了严重的预测失误
 - 大数据自大说明，数据不能替代常识或严谨的研究设计
- **经济学启示：**在数据经济时代，**预测能力**本身就具有极高的商业价值，推动了研究范式从传统的假设驱动向数据驱动转变

对经济学研究方法的拓展

传统计量经济学 vs. 大数据分析

计量经济学

- **数据类型：**
 - 结构化数据（截面、时间序列、面板），有清晰的经济学含义
- **核心理念：**
 - 假设检验（证伪主义），证实或证伪某个经济理论

大数据分析

- **数据类型：**
 - 包含非结构化数据（文档、视频、图像），难以用通常形式可视化
- **核心理念：**
 - **预测**扮演更重要角色，关注预测效果的评估

- **面临的挑战：**

- 需要访问数据、处理大规模数据集的编程能力
- 需要创造性的方法来总结及分析信息

- **重要提醒：**

- 数据不能替代常识、经济理论或严谨的研究设计

- **未来展望：** 数据推动的经济学研究方法和范式转型还需要更多深入研究

第三节

全球视域下数据经济学的 核心研究议题

数据经济学定义

一门试图解释在数据要素参与条件下，产品和服务的生产、交换、分配、消费等经济运行全过程的经济学。

- 以计算机科学、法学、管理学、经济学等多学科为基础的交叉科学
- 三大核心议题：
 - ① 数据产权
 - ② 数据监管
 - ③ 数据隐私

数据产权：是否应该有所有权？

- **反对派观点**（德国马克斯·普朗克创新与竞争研究所）：

- 所有权会形成壁垒，阻碍数据的可获得性
- 相当于授予专有垄断权，会导致**数据市场的扭曲**

- **主流观点：**

- 数据**应该**被赋予所有权
- 争论焦点转移至：是单一产权还是多重产权？产权赋予谁？

数据产权：权利结构设计

1. 单一产权思路

- 基于数据的有限排他性
- 创设**数据生产者权**
- 类似于著作权、专利权，通过专门立法保障

2. 协同格局思路

- 双重所有权（名义 vs. 实际）容易制造权利纷争
- 改进方案：**所有权 + 用益权**
 - **数据原发者**：拥有数据所有权
 - **数据处理者**：拥有数据用益权

数据产权：产权应归属于谁？

Jones & Tonetti (2020) 理论模型发现：

基于数据是非竞争性的假设，不同归属会导致不同后果：

① 公司拥有时：

- 出于利益驱动，可能**过度使用**数据（不尊重隐私）
- 出于对“创造性破坏”的担心，不愿出售数据（**囤积**），导致福利损失

② 政府限制出售时：

- 出于保护隐私完全禁止出售，行为**特别有害**

③ 消费者拥有时：

- 可以实现**接近最优**的数据分配

案例解析：消费者拥有数据产权的实践

场景：欧盟通用数据保护条例（GDPR）与数据可携权

赋予消费者数据产权可以打破企业的垄断，并促进最优分配。

- **企业囤积的弊端：**

- 社交平台或医疗机构倾向于将用户数据视为核心资产并拒绝共享
- 导致新创企业无法获取训练数据，阻碍了市场的创造性破坏

- **数据可携权的经济学意义：**

- 法律赋予用户随时将个人数据打包带走并转移给竞争对手的权利
- 消费者通过用脚投票，迫使企业提供更优质的服务，实现数据的接近最优分配

案例解析：从隐私保护到数据价值流通的立法演进

场景：欧盟从 GDPR 向《数据法案》的演进

随着物联网的发展，数字经济的前沿争议正从“个人隐私”拓展向“非个人工业数据的确权与流通”。

- **早期重心 (GDPR)**：侧重于个人数据的**隐私保护**与消费者防卫
- **最新前沿 (数据法案)**：
 - **破除数据孤岛**：针对智能工厂、网联汽车等物联网设备产生的大量非个人工业数据，打破设备制造商的绝对垄断
 - **强制共享与用益权**：赋予用户（B 端或 C 端）访问和共享设备生成数据的权利，并允许将其转移给第三方维修或创新服务商
- **经济学意义**：
 - 标志着监管重心从防御性保护转向**促进数据要素流通与价值创造**
 - 通过法律界定数据原发者与处理者的**数据用益权**，激活了售后服务和第三方创新的二级市场

数据监管：核心问题与争议

核心问题： 1. 在何种情况下需要监管？ 2. 什么样的政策能实现目标？

观点： 监管可能是有害的

过度的隐私保护会扭曲市场竞争，阻碍创新。

- **成本逻辑：** 政府隐私保护的**成本往往远高于**侵犯隐私带来的成本
- **实证证据：**
 - **美国信用卡行业：** 严格隐私法阻碍了甄别，导致高抵押价格、高违约率
 - **欧盟在线广告：** 隐私保护政策明显降低了广告市场有效性，对**小规模平台**负面影响更突出

数据监管：最优制度设计的思路

思路一：强调私人规制

- 机制：声誉机制、增加信任
- 发现：
 - 政府强制标准往往不能带来社会福利最大化
 - 商家自律（买家自行注意）和第三方认证能提高福利

思路二：综合运用多种机制

- 框架：共同监管（Co-regulation）
- 关系：互补而非替代
- 政府的角色：
 - 结合市场、私人规制和技术方案
 - 只有在其他机制失效且收益明显高于成本时，政府监管才是必要的

案例解析：私人规制与共同监管的发挥

场景：第三方隐私认证与平台隐私标签

政府强制监管成本极高，市场化与行业自律往往能提供更高效的解决方案。

- 私人规制的有效实践：

- **第三方认证**：网站悬挂 TRUSTe 等隐私印章，由独立机构监督其隐私政策执行，有效增加消费者信任
- **平台自律**：苹果应用商店强制要求开发者披露隐私应用标签，让用户在下载前清晰知晓数据收集情况

- 经济学启示：

- 依靠声誉机制和买家自行注意，可以避免一刀切式政府监管对中小企业造成的创新阻碍

数据隐私：成本与影响

- 隐私成本的构成：

- ① 直接成本：个人对私密性的主观偏好 + 直接经济损失
- ② 间接成本：由隐私诱发的策略性行为（例如：卖家利用信息进行价格歧视）

- 影响的差异性：

- 影响大：健康信息、社交媒体信息
- 影响小：语音识别数据、驾驶数据（前提是充分匿名且无不利影响）

案例解析：隐私成本异质性与适度保护

场景对比：社交数据滥用与交通数据共享

不同类型的数据对个人生活安宁的影响存在显著差异。

- **高隐私成本场景：剑桥分析事件**

- 收集海量用户的性格特征与社交偏好，引发精准的政治广告推送
- 极大地侵犯了主观私密性，甚至诱发了策略性操纵，必须严格保护

- **低隐私成本场景：城市交通拥堵治理**

- 导航软件收集大量充分匿名的车辆行驶速度与轨迹数据
- 对个人无明显不利影响，且能产生巨大的公共利益，应鼓励共享

数据隐私：两种研究视角

法学视角：隐私人格权

- **定义：**隐私是一项基本人权（自由与安静）
- **核心：**
 - 确立私人空间和公共空间的边界
 - 保证个人对数据的**自决权**
 - 增强个人对隐私的**控制**

经济学视角：隐私经济品属性

- **定义：**个人数据是一种重要的生产要素
- **核心：**
 - 关注权衡：保护隐私 vs. 促进数据共享
 - 追求**适度保护**，而非绝对保护
 - 目标是效率或社会总福利最大化

数据隐私：经济学研究的重点

基本认识

市场化的个人数据交易机制可以同时实现有效的隐私保护和数据开发利用。

- **研究焦点：**
 - **确定边界：** 确定个人信息封闭和信息披露之间的**最优边界**
 - **激励机制：** 探讨经济激励机制在促进最优隐私保护中的独特作用

本章总结一：核心概念

- 数据的要素化：

- 数据已不再是单纯的信息载体，而是核心的生产要素

- 对传统经济学的冲击与拓展：

- 增长理论：数据驱动经济成为内生增长的特例，机器知识资本成为新引擎
- 信息不对称：在降低传统搜索成本的同时，催生了基于算法模型的新型不对称
- 交易成本：大幅削减交易成本，促使企业物理边界消融与生产网络化

本章总结二：研究范式与现实治理

- 研究范式的第四次革命：

- 经济学实证研究正在经历巨变，研究重心从传统的追求因果关系转向拥抱混杂数据的**相关性预测**

- 数据经济的三大治理基石：

- **数据产权**：清晰的产权是要素分配的前提，赋予消费者数据产权有助于打破平台囤积，实现资源接近最优分配
- **数据监管**：单一的政府强制干预容易扼杀创新，市场在考虑私人规制与行业自律相结合的共同监管体系
- **数据隐私**：隐私具有经济品属性，核心诉求是在个人生活安宁与数据共享红利之间，寻找到实现社会总福利最大化的最优边界

谢谢大家！