

数据经济学

第四章：数据的需求

陈希路

暨南大学经济学院

2026 年春

- 1 第一节：数据的应用场景：产业链视角
- 2 第二节：数据的买方异质性
- 3 第三节：数据需求侧的收益函数

章首导言：数据需求的主体与视角

- 数据需求方的分类：

- 消费者（消耗）：通过消费数据产品或服务获取自身效用提升
- 企业（再生产）：将数据作为生产要素，创造各类产品与服务

- 引例：短视频成瘾背后的经济逻辑

- 现象：用户投入时间越多，推送越合乎口味
- 机制：数据与算法结合 → 精准匹配
- 结果：流量带来强网络效应与广告收入，最大化平台收益

第一节

数据的应用场景：产业链视角

数据的应用场景：产业链视角

从产业链视角来看，数据在以下三个环节中起到至关重要的作用：

- ① 研发创新
- ② 生产制造
- ③ 市场匹配

创新的本质 (Agrawal et al., 2019)

在高度复杂的知识空间中发现既有知识的全新组合。

- 知识积累的双刃剑：
 - 初期：有助于排列组合形成新知识
 - 后期：知识空间高度膨胀，寻找有用组合如大海捞针，成为创新的负担
- 数据的贡献：
 - 驱动人工智能技术，预测具有最高发展潜力的组合
 - 提高创新发现率

数据在研发创新中的应用：失败数据的经济学价值

试错成本与信息外部性

创新本质上是试错的过程，失败尝试所形成的数据对行业整体具有避坑价值。

- **赢家通吃下的双寡头模型 (Akcigit & Liu, 2016):**
 - 假设两家竞争企业起初均采用**高利润/高风险**的研发路线
 - 探索过程中，某家企业可能发觉当前路线是死胡同
- **企业的占优策略:**
 - 隐瞒失败数据，秘密转向安全但低回报的渐进式研发路线
 - 动机：防止竞争对手获取信息进而少走弯路

数据在研发创新中的应用：分散均衡下的效率损失

- 缺乏数据共享的后果：

- 无谓的重复投资：另一家企业大概率将时间和资源浪费在已证实无效的项目上
- 全社会的创新停滞：因观察不到对手进展，企业可能因不确定性过高而集体提前放弃颠覆式创新

- 结论：

- 数据的私有化保护导致了严重的信息外部性缺失
- 建立行业级的基础研发数据共享机制（或数据要素池）是消除无谓损失的关键

案例分析：数据在研发创新中的应用

英矽智能人工智能药物研发

利用海量生物学数据与人工智能算法，缩短新药研发周期、降低试错成本。

- **传统困境：** 寻找有效药物分子靶点如同大海捞针，且失败率极高
- **数据赋能：**
 - 整合数百万组学数据样本与文本数据
 - 训练生成式人工智能模型，快速预测最具潜力的化合物结构
- **成效体现：** 将靶点发现到候选药物确定的时间从数年缩短至几个月

数据在生产制造中的应用

- **核心作用**：提高生产、决策和管理效率
- **具体功能**：
 - **优化流程**：分析数据发现瓶颈
 - **质量控制**：传感器实时监测关键参数（温度、湿度、压力等）
 - **安全保障**：监测设备风险，及时维护
- **积累公式**：数据要素像资本一样积累：

$$D_{t+1} = I_t + (1 - \delta)D_t$$

(D ：资本积累， I ：投资额， δ ：折旧率)

数据要素的两重创新性特征

- ① **直接提升生产效率**：如改进自动驾驶算法
 - ② **提升自身积累效率**：通过循环迭代，提升分析处理能力
-
- **实证证据 (Müller et al., 2018)**:
 - 拥有大数据分析资产使企业生产率平均提高 **4.1%**
 - 信息技术密集型行业更是高达 **6.7%**

3D 技术：数据驱动型决策 (Data-Driven Decision Making, DDD)

- **定义：** 以数据及从中挖掘出的模式和洞见进行决策，减少对直觉的依赖
- **成效：**
 - **Brynjolfsson et al. (2011)：** 为美国企业贡献了 5%–6% 的产出和生产力增长
 - **McAfee & Brynjolfsson (2012)：** DDD 占比最高的企业，平均生产效率高出 5%，利润率高出 6%

案例分析：数据在生产制造中的应用

海尔卡奥斯工业互联网平台

将生产制造环节数字化，实现数据驱动型大规模定制。

- **应用场景：**打通用户交互到生产交付的全流程数据
- **数据驱动：**
 - 实时收集车间设备运行状态，进行预测性维护
 - 结合供应链数据，系统自动匹配最优生产排期与物料调度
- **经济效益：**显著降低库存积压，提高劳动生产率并入选全球灯塔工厂

数据在市场匹配中的应用

现象：大数据杀熟

企业利用数据将用户分类，进行价格和服务歧视，如不同手机品牌用户打车费用不同。

- 微观经济学视角的解释：

- 传统市场中，企业通常只能实施**三级价格歧视**（按粗略群体定价）
- 细颗粒度的数据使得企业具备了向**一级价格歧视**（**完全价格歧视**）逼近的能力

- 福利分配效应：

- 平台利用算法精准测算出每位消费者的**保留价格**（**支付意愿**）
- 结果：消费者剩余被极大地转化为生产者剩余，最大化了平台收益

数据在市场匹配中的应用

- 事前应用：

- 了解用户画像（人口特征、偏好）
- 案例：社交软件数据辅助快餐店选址；保险公司评估风险等级进行差异化定价

- 事后应用：

- 分析反馈数据，优化产品设计
- 案例：宝马分析试驾数据消除隐患；IBM 缩短产品优化周期

- 结合 AI 的高级应用：

- 案例：汽车 4S 店利用语音识别采集销售谈话数据
- 效果：自动实时评分，提高营销环节的管理和服务效率

案例分析：数据在市场匹配中的应用

希音 Shein 敏捷供应链与精准营销

通过海量消费数据捕捉前端流行趋势，并实时匹配后端生产。

- 事前洞察：

- 爬取分析全球社交媒体与搜索平台数据
- 预判流行元素，精准描绘目标客群画像

- 事后优化：

- 每日上新并实时追踪商品点击率与加购率
- 销售数据表现好的款式立刻返单生产，表现差的迅速淘汰

第二节

数据的买方异质性

数据的买方异质性

- **现象：**数据本身在来源、使用目的、规模、质量等方面具有显著差异
- **核心观点：**
 - 不同种类的数据满足不同的需求
 - 同样一份数据在不同的企业中能够创造的价值不尽相同
- **结果：**这种异质性直接表现在**数据价格**上

数据需求的性质差异

(一) 实时数据 vs. 历史数据

实时数据：最新最快的数据

在竞争激烈的环境中，取得优势的机会稍纵即逝，实时数据价值最高。

典型应用场景：

- **金融业：**
 - 投资：监控市场动态，及时调整策略
 - 监管：识别风险行为，防欺诈洗钱
- **交通运输业：**车路协同、事故预警
- **公共部门：**政策效果评估、实时反馈

数据需求的性质差异

(一) 实时数据 vs. 历史数据

实时数据在金融高频交易（HFT）中的价值

- **高频交易的本质**：利用极短时间内的市场微观结构变化与微小价差获利
- **实时数据作为核心生产要素**：
 - **数据维度**：毫秒级甚至微秒级的订单簿数据、逐笔成交数据
 - **基础设施竞赛**：金融机构耗巨资铺设专用光缆或微波基站（如芝加哥到纽约专线），仅为减少几毫秒传输延迟
- **启示**：
 - 数据的价值随时间呈现**指数级衰减**，上一秒的数据在 HFT 中价值归零
 - 实时数据的买方异质性极高，对量化机构价值高，对长期投资者则价值低

数据需求的性质差异

(一) 实时数据 vs. 历史数据

历史数据

利用不同时间节点收集的数据，反映事物随时间变化的状态或程度。

典型应用场景：

- 科研机构以及金融、政府部门：验证因果关系、预测趋势（时间序列分析）
- 互联网平台（如网飞 Netflix）：
 - 机制：积累用户行为（点击、播放、评分）→ 完善用户画像
 - 价值：精准推荐内容，形成难以匹敌的**竞争优势**

数据需求的性质差异

(二) 专有数据 vs. 公共数据

专有数据

一家企业单独所有，并可以阻止其他企业使用的数据。

- **重要性：**是 AI 相关企业发展最重要的投入
- **数据反馈循环：**
 - 专有数据 → 训练独有算法 → 差异化产品 → 吸引更多用户 → 生成更多数据
- **结果：**建立天然壁垒，产品可替代性低

案例分析：专有数据的竞争优势

科技巨头的大语言模型之争

专有数据成为构建顶级人工智能模型的核心护城河。

- **公共数据瓶颈：** 各家企业均可轻易获取维基百科等开源数据集，难以形成差异化
- **专有数据价值：**
 - 顶级科技公司通过海量活跃用户建立极其庞大的专有对话与反馈数据库
 - 持续利用人工反馈强化学习机制优化模型表现
- **市场结果：** 独家数据飞轮效应使得后发者难以超越，确立了领先垄断优势

专有数据与企业竞争优势：数据反馈循环

理论基础 (Farboodi & Veldkamp, 2020)

利用专有数据训练独有算法，会形成数据反馈循环，这是人工智能企业确立优势的核心。

- 循环机制：

- 投入专有数据训练模型，开发出具备差异化和低可替代性的人工智能产品
- 更优质的产品吸引更多消费者使用
- 消费者的使用行为进一步生成更多全新的用户数据，反哺下一次模型迭代

- 经济学结果：

- 相比仅利用公共数据的企业，拥有数据反馈循环的企业能创造出需求弹性较小的产品
- 提高了竞争对手的模仿门槛，在市场中确立垄断地位

数据需求的性质差异

(二) 专有数据 vs. 公共数据

公共数据

政府发布的大型数据集（如统计局数据）。

- **科研用途**：实证研究的基础（涵盖各级地区、时间维度）
- **现实需求（智慧城市）**：
 - **政府**：识别问题，感知城市状况
 - **企业**：开发应用（如地图导航、预警）
 - **消费者**：掌握服务状态，获取信息

案例分析：公共数据与智慧城市生态

2012 年美国纽约市《开放数据法案》

通过立法强制开放政府数据，赋能多方主体，打造智慧城市生态系统。

- **政策背景：**到 2018 年，除安全与隐私数据外，纽约市政府及分支机构数据须依法向公众开放
- **赋能效果：**
 - **政府治理：**基于数据识别城市管理短板，实时感知交通、卫生和治安状况
 - **企业商业化：**开发交互式地图、犯罪预警应用及最优线路规划服务
 - **消费者福利：**无缝获取城市服务状态信息，提升生活便利度与决策效率

相同数据的需求异质性

为什么数据很难标注统一的“建议零售价”？

- **非标准化**：潜藏的信息往往包含丰富的行业专门知识
- **价值差异大**：
 - 同样一份数据在不同需求方处的价值差异显著
 - 若无法创造价值，保留数据甚至是“有毒”的（存储成本高）
- **定价策略**：数据提供商往往通过谈判、拍卖或垄断定价等形式实行**价格歧视**

相同数据的需求异质性

案例：英超球员数据卖给谁？

需求方	支付价格	身份	用途
需求方 A	10 万英镑	俱乐部数据分析团队	“刚需”，用于制定战术赢取积分
需求方 B	5000 英镑	狂热球迷	“得力助手”，用于社区讨论
需求方 C	0 元	咖啡店老板	“累赘”，毫无兴趣，不愿占用硬盘

启示：价格差异反映了同一份数据对不同需求方的价值差异

相同数据的需求异质性

数据提供商制定价格策略时需考虑的因素：

❶ 数据质量：

- 精度更高、更新频率更快的数据通常价格更高

❷ 买方购买力：

- 某些买方（如职业俱乐部）可以承担更高价格

❸ 需求范围：

- 部分数据 vs. 全部数据
- 例子：球迷只买主队数据 vs. 球队需要全联盟对手数据

第三节

数据需求侧的收益函数

数据需求侧的收益函数

数据需求方主要分为两种角色，其收益来源不同：

1. 企业（生产要素）

- **用途**：用于生产各类产品和服务
- **收益**：为企业创造商业收益

2. 消费者（产品/服务）

- **用途**：满足消费者需求
- **收益**：体现为消费者的**效用**（从产品中获得的**价值**）

企业收益函数：基本逻辑

- **基本公式**：企业的收益 R 可表示为 $R = P \cdot Y$ ，其中 P 为价格， Y 为产量
- **核心问题**：
 - 在完全竞争市场中，企业是价格 P 的接受者
 - 因此，关键在于数据要素如何影响产量 Y
- **理论挑战**：
 - 传统的经济学模型并未包含数据
 - 大多数研究将数据加入知识生产环节，促进新知识产生

方式一：数据作为独立的生产要素

- 柯布-道格拉斯生产函数形式：

$$Y = AK^{\alpha}D^{\beta}L^{1-\alpha-\beta}$$

- 符号含义：

- A ：技术水平； K ：资本； L ：劳动力
- D ：数据要素
- α, β ：各要素产出在总产出中的份额

- 特点：

- 数据要素与资本、劳动力并列，**独立**发挥作用
- **关键性质**：若数据投入为零，整个经济系统产出为零

方式二：数据作为资本的互补（或替代）要素

- 生产函数形式：

$$Y = A(\mu K^\rho + D^\rho)^{\frac{\alpha}{\rho}} L^{1-\alpha}$$

- 特点：

- 数据要素 D 与资本 K 密切相关
- 根据参数 ρ 的不同，存在**互补**或**替代**关系

- 直观理解：

- 例如：**数据 + 算力 = 最大化价值**
- “巧妇难为无米之炊”：算力不足影响效率，数据不足影响产出

- 与方式一的区别：

- 若数据投入为零，可能会影响产量，但**不至于使得产出归零**

方式三：数据作为研发的投入要素

- 知识生产函数：

$$\dot{A} = f(D, A, L_A)$$

- 符号含义：

- $\dot{A}$ ：新知识（技术）的生产
- A ：知识存量； L_A ：投入知识生产的劳动力
- D ：数据要素投入

- 核心逻辑：

- 数据是能够提升想法或知识质量的动力

- 案例：特斯拉自动驾驶

- 少量数据 → 仅支持紧急制动
- 大量数据训练 → 机器学习算法升级 → 实现复杂路况自动驾驶

知识回顾：数据要素引入生产函数的宏观增长含义

将数据引入生产体系，彻底打破了传统经济学的部分假设边界：

① 边际收益可实现递增：

- 传统资本和劳动力存在边际收益递减
- 数据要素具**非竞争性**，多主体反复使用不损耗，反可通过交叉验证提升数据质量

② 规模报酬递增：

- 数据的零边际复制成本结合网络效应，使得总生产函数呈现规模报酬递增
- 数据富集型企业或国家容易形成“强者恒强”的内生增长路径

消费者效用函数：效用与隐私的权衡

参考 Jones & Tonetti (2020) 模型，数据对消费者有两方面影响：

1. 正向效用

- 数字产品或服务像普通商品一样带来满足感

2. 负向效用

- 企业利用数据进行价格歧视、推送骚扰广告等
- 侵害权益，带来隐私成本

消费者效用函数模型

假设消费者在时期 t 的效用函数为：

$$u(c_t, x_{it}, \tilde{x}_{it}) = \underbrace{\int_0^{N_t} \log c_{it} di}_{\text{消费带来的正效用}} - \underbrace{\frac{\kappa}{2} \frac{1}{N_t^2} \int_0^{N_t} x_{it}^2 di}_{\text{企业自用数据的负效用}} - \underbrace{\frac{\tilde{\kappa}}{2} \frac{1}{N_t} \int_0^{N_t} \tilde{x}_{it}^2 di}_{\text{数据共享带来的负效用}}$$

变量定义：

- c_{it} ：个体的消费（产生数据）
- x_{it} ：生产商品的企业实际用于生产的数据比重（如特斯拉自用）
- $\tilde{x}_{it}$ ：其他企业共享的数据占比（如特斯拉数据被 Waymo/GM 使用）
- $\kappa, \tilde{\kappa}$ ：隐私敏感度参数，消费者越敏感，负效用越大

消费者的核心权衡与隐私悖论

- **效用的净值考量：**

- 个体能从数字化服务中获得便捷性与定制化正向效用
- 但也面临被算法剥削、数据泄露带来的严重负面成本
- 最终效用水平是上述两者的净结果

- **隐私悖论现象：**

- 问卷调查中，消费者普遍声称高度重视个人隐私数据
- 现实行为中，消费者却往往为了微小的便利（如免注册、小额优惠券）轻易让渡核心数据隐私
- 根源：由于信息不对称与隐私条款的晦涩，消费者无法准确评估公式中的负效用项 κ

知识回顾：数据产权归谁才能实现福利最大化？

Jones & Tonetti (2020) 模型的政策含义

数据要素的非竞争性决定了数据被越广泛地共享，社会创造的总效用越大。

● 情形一：数据产权归企业所有（现状）

- 企业有囤积数据的动机（防止竞争对手获益），导致数据成为孤岛
- 结果：扼杀创新，且企业倾向于过度使用消费者数据，导致社会福利受损

● 情形二：数据产权归消费者所有（最优解）

- 消费者权衡隐私成本后，可主动将数据授权给能提供高质量服务的企业
- 结果：数据实现了**最大范围的共享**，同时尊重了隐私偏好，达到了经济学意义上的帕累托最优

本章总结：应用场景与需求异质性

- **数据赋能产业链全流程：**

- 研发端加速知识重组与创新发现速率
- 生产端依靠数据驱动型决策持续优化制造流程
- 市场端实现供需精准匹配与前端动态反馈

- **需求主体的买方异质性：**

- 时效差异划分出实时数据与历史数据的不同应用空间
- 专有数据成为企业建立竞争壁垒与飞轮效应的核心关键投入

- **非标准定价逻辑：**

- 同一数据源对不同买家的潜在价值大相径庭
- 供给方据此采取价格歧视策略以实现销售收益最大化

本章总结：收益函数与核心权衡

- **重塑企业生产函数：**

- 数据作为全新要素直接进入模型并影响产出规模
- 数据亦可作为资本互补品或研发催化剂进而提升技术上限

- **消费者效用博弈：**

- 个体享受海量数字化服务带来的正向福利提升
- 用户同时必须承受个人信息被企业自用或共享产生的隐私隐患

- **整体经济学启示：**

- 数据要素极大拓展了传统经济学理论对生产要素的定义边界
- 合理释放数据价值并平衡隐私保护是实现数字经济增长的必由之路

谢谢大家!