罗汉堂
Luohan Academy

Luohan Academy

理解大数据：
数字时代的数据和隐私

2021

作者包括：

帕特里克·伯顿
哥伦比亚大学，商务学教授，
美国文理科学院院士

陈龙
罗汉堂，总裁

本特·霍斯特罗姆
麻省理工学院，经济学教授，
诺奖得主

埃里克·马斯金
哈佛大学，经济学教授，
诺奖得主

克里斯托弗·皮萨里德斯
伦敦政治经济学院，经济学教授，
诺奖得主

迈克尔·斯宾赛
斯坦福大学，经济学教授，
诺奖得主

孙涛
国际货币基金组织，高级经济学家

孙天澍
南加州大学，工商管理学教授

熊伟
普林斯顿大学，金融学教授

杨立岩
多伦多大学，金融学教授

每日免费获取报告

1、每日微信群内分享**7+**最新重磅报告；

2、每日分享当日**华尔街日报**、金融时报；

3、每周分享**经济学人**

4、行研报告均为公开版，权利归原作者所有，起点财经仅分发做内部学习。

扫一扫二维码
关注公号
回复：**研究报告**
加入“起点财经”微信群。。

数字时代的数据和隐私

2021 年 6 月 2 日

作者名单

罗汉堂社区	罗汉堂内部
帕特里克 · 伯顿 (Patrick Bolton)	陈龙
本特 · 霍斯特罗姆 (Bengt Holmström)	黄亚东
埃里克 · 马斯金 (Eric Maskin)	李勇
克里斯托弗 · 皮萨里德斯 (Christopher Pissarides)	罗璇
迈克尔 · 斯宾赛 (Michael Spence)	马英举
孙涛	欧阳书淼
孙天澍	朱锋
熊伟	
杨立岩	

* 感谢以下学者和专家为报告提供的真知灼见和批评指正：Nageeb Ali, Benjamin Bai, Xiao Cheng, James Dempsey, Nico van Eijk, Isabelle Falque-Pierrotin, Fuping Gao, Clarisse Girot, Jane Horvath, Rohit Lamba, Haiying Li, Weiqiu Long, Wei Lv, Ariane Mole, Zhengjun Nie, Pietro Ortoleva, Thomas Sargent, Jean Tirole, Laura Veldkamp, Neng Wang, Stephen Kai-Yi Wong, Hanhua Zhou, 以及 2019 杭州隐私和数据治理研讨会、罗汉堂第二届年会、2020 年北京数据与隐私研讨会、2020 年罗汉堂-普林斯顿大学 Bendheim 金融研究中心数据与隐私线上研讨会的所有参与者。此外我们在此感谢 Thomas Walton 和 Michael Whinihan 在报告写作和编辑工作中提供的宝贵建议。

序言

人类早已意识到，社会和经济的正常运转，以及对经济福利的追求都离不开信息的处理和分享。例如，在真实世界中从事生产活动，我们必须获取并分享周遭环境的信息；要进行社交，我们也要获取并分享他人的信息。“了解你的客户”，从而为他们提供高质量产品和服务是今天取得商业成功的关键。黄页电话簿的诞生和广泛使用，说明个人向公众分享个人信息，加强与社会联结，已经成为现代人际关系的基础。

进入 21 世纪，数据信息的使用如此盛行，以至于我们称今天的时代为“大数据时代”。数字信息让过去难以实现的社会协作成为可能，大大提升了公众福利，但也引发了人们对数据和隐私问题的焦虑：**我们如何在大数据时代保护个人隐私？数据使用创造的价值归谁所有，该如何分配？如何理解数据使用产生的风险？大数据应用是否会带来“赢者通吃”的市场现象，从而妨碍竞争，损害消费者和整个社会的利益？**这些正是本报告尝试回答的问题。

我们试图“用大数据研究大数据”，基于尽量客观的实证证据，评估大数据对社会经济带来的影响。如同诺奖获得者科斯（1994）所言，我们需要远离纯粹的“黑板经济学”，因为那只存在于理论家的头脑之中。他说：“我们需要更多的实证工作……一位充满灵感的理论家也许不需要，但是……这些灵感大多受到现象中的规律、悖论或异常现象的启发，而这些都依赖于系统的数据收集，尤其是，当我们的首要任务是打破固有的思维定式时。”

科斯的这一提醒，对研究数据问题尤为重要。**数据和传统的生产资料不同，具有非竞争性和与经济活动、数据相关主体的不可分离性**。需要基于对数据本质的理解，破除原有的基于理解传统的有形商品和要素的思维定式，探索基于实证的、整体性的、符合多方利益的方法，来保护数据和隐私，否则虽以保护为名，无意中却难免损害公共利益，这无异于因噎废食，并将错过数字技术带来的重大机遇。

是为序。

罗汉堂

目录

第 1 章 报告综述

“经济社会的首要问题，是利用好分散在个人手中的信息的问题。这是因为我们在决策场景中所必需的知识，从来不是以整体的方式存在，而是以不完整、甚至经常矛盾的方式散落在不同个体手中。”

--哈耶克，1945

1.1. 一个由数字信息定义的时代

人类的经济发展史是一部信息分享的历史。和别的动物不同，千年以来，人类学会了收集、组织和储存大量复杂信息，并彼此分享。然而大千世界中，永远有人类难以收集的信息，或是缺失，或是不够精确，更不用说对这些信息进行处理并从中受益了。更复杂的，由于个人和企业是自利的，或有意或无意地，他们常常提供错误或不完整的信息。

为何信息分享如此重要？**哈耶克对此有深刻的认识，他认为：“经济社会的首要问题，是利用好分散在个人手中的信息的问题。这是因为我们在决策场景中所必需的知识，从来不是以整体的方式存在，而是以不完整、甚至经常矛盾的方式散落在不同个体手中。”**（哈耶克，1945）。他相信，推进信息分享是经济社会最重要的问题，攸关人类的经济福祉。

在理论界，过去的大半个世纪中，众多经济学家致力于研究信息的价值，并探索如何降低乃至消除信息分享的壁垒。1993 年诺奖得主 Douglas North 认为：“协作的根本理论问题在于个人如何探知他人的偏好和行为模式”（North，1990）。在实践中，生产者越了解客户，就越能更好地服务客户。在美国，一百多年来，在每一个城市和小镇，每户家庭的基本个人信息，包括姓名、住址和电话，通过黄页的形式，都可以公开查阅到，其目的是为了帮助社会成员找到对方，促进个体与社会的连接。在医药和金融等领域，“了解你的客户”（knowing your customer，KYC）是用户获得高质量服务的前提，其中常常涉及隐私和敏感的个人信息。

信息之所以需要分享，还因为信息存在“不对称”的问题。在相互接触中，人们获得的常常是不同类型的信息。而双方往往不愿或无法可信有效地交换信息，因此，“不对称”的信息很难恢复“对称”。而信息不对称会降低经济效率，是影响市场交易效率的重要因素之一（Spence，1973，1974; Grossman and Stiglitz，1980）；甚至，当信息不对称严重到一定程度，整个市场可能会就此消失（Akerlof，1970）。例如在劳动市场，缺乏有关工人能力以及企业用工需求的信息，会导致人力资源无法得到有效分配，现实表现就是失业和企业生产率低下（Phelps，1970; Pissarides，2000）。

这些思考让另外一位诺奖得主科斯推断，大部分经济活动，其背后的机制设计，本质上都是“为了降低交易成本，或弥补过高交易成本引发的交易失败，从而让个人可以

自由地协商交易，并如哈耶克所言，受益于信息的扩散”。一方面，为了促进信息的收集和扩散，人类进行了诸多努力，设计不同的机制来减少交易成本。另一方面，当经济主体面临不充分或不对称信息时，我们设置不同的激励措施来鼓励主体之间进行协作（例如 Hart，1988; Hart and Moore，1988; 以及 Holmström，1979，1982）。所以，在理论和实践中，人类社会都一直在致力于打破信息的牢笼，推进信息的交互。

20 世纪 40 年代，克劳德 · 香农和阿兰 · 图灵的天才创想，将数据编码于“数字原子”中——今天被称为比特。自此，数据可以被数字化，现代信息科学就此滥觞。再加上新兴的半导体技术被大规模用于数据计算和储存，推动了数据的爆炸式增长。以至于到 20 世纪 70 年代，在各类文本中，“数据”一词的出现频率超越了“信息”一词。数字革命彻底改变了信息在社会和经济中扮演的角色。它一方面让信息获取前所未有地简单，人类使用信息并从中获益的难度大大降低，数据日益成为重要的生产要素。另一方面也加大了滥用信息的风险。

因为数据、信息、大数据等概念被频繁使用，在进一步论述之前，有必要建立对这些概念的共识。首先要注意的是，数据不等同于信息。数据是对事物的一系列观察，而“大数据”则是对大量“(小) 数据”进行组合、存储和计算处理的过程。“数据科学”中很重要的步骤是“数据压缩”，即将大量数据集压缩成小规模，同时保留大部分有效信息的数据集，并将其转换成易于存储和解读的形式。信息是基于数据的洞见，所以信息的价值往往取决于需要回答什么问题，也就和具体的使用场景相关。

数据的广泛使用，引发了人们对三个问题的思考：**我们如何在大数据时代保护个人隐私？数据应该归谁所有，以及该如何分配数据使用产生的福利和风险？大数据应用在多大程度上会带来“赢者通吃”的现象，从而阻止竞争，损害消费者和整个社会的利益？**

这些是我们在本报告中尝试回答的主要问题。人类社会正处于一个关键的十字路口，一方面数据的重要性达到前所未有的高度，另一方面我们在如何治理，从而充分发挥数据的作用，并同时降低数据滥用的风险这些关键问题上，鲜有共识。要发挥数据的最大效用，需要我们更好地理解数据的本质，理解数据在真实世界是如何使用的，以及在治理数据时应该如何权衡取舍。为此，我们要区分事实和臆断、求真与恐惧。**在数据时代更好地理解数据的经济学本质，或者说“数据经济学”正是本报告的主要课题。**

数据经济学就是数字时代的信息经济学。本报告一方面汇聚了一批合著者，他们深刻理解信息的本质和价值，并因为对其的研究获得诺贝尔经济学奖；另一方面聚焦对“大数据”应用的详实证据，尝试用事实说话，评估“大数据”的真实影响。面对这样一个重要的话题，尊重事实而非臆测尤其重要。如同科斯（1994）所言，要远离纯粹的“黑板经济学”。他说：“我们需要更多的实证工作……一位充满灵感的理论家也许不需要，但是……灵感大多受到现象中的规律、悖论或异常现象的启发，而这些都依赖于系统的数据收集，尤其是，当我们的首要任务是打破固有的思维定式时。”

1.2. 理解数据的本质

1.2.1 从理解隐私悖论开始

要保护好隐私，我们需要了解人们在真实生活中如何看待和做有关个人数据分享的决策。

我们可以先看一个简单的、被广为接受的关于信息或数据隐私的定义。美国最高法院大法官 Louis Brandeis 认为："对个人信息的可控性，是个人应享有的'独处的权利'（to be left alone）"（Pavlou，2011）。从这个角度出发，我们进一步追问，当人们为了享受数字服务的好处，必须提供一定个人信息时，他们是如何为自己"独处的权利"进行决策的？在回答这个问题时，大量研究发现，全球用户中普遍存在着一种矛盾现象，学者们称之为"隐私悖论"，它描述的是，尽管大多数人表示在意自己的隐私，但常常免费地，或在很小的经济补偿下，分享自己的个人信息。人们对隐私基本权利的重视，和他们实际行为中的"毫不在意"之间存在显著矛盾。这种现象并非孤例，存在于不同的国家和文化环境中。

目前学术界对隐私悖论有几种不同的解释。一种观点认为，这是因为当事人不了解隐私被侵犯可能带来的严重后果；或由于一些重要的数字应用缺乏可替代选择，用户不得不让渡一定的隐私权（Chen and Michael，2012）。例如，为了使用微信，用户必须同意微信的隐私条款，否则只能退出。但随着越来越多的新选择不断涌现，这种理论很难解释用户为何对层出不穷的新数字服务也"来者不拒"。另一种更让人信服的解释认为，当面临真实的选择时，是人们的真实行为，而不是调研中的表达，揭示了人们会在隐私和数据福利之间如何权衡取舍的真相。

因此问题的关键在于，当用户在市场中真正拥有选择权时，他们是如何决策的。为了回答这个问题，在下文第二章中，利用支付宝数据，我们进行了一项大规模实证分析。支付宝活跃用户众多（超过 10 亿），且有大量小程序可以选择。这些小程序有的来自小型初创企业，有的来自成熟大企业。使用小程序时，企业需要获得用户许可来获取一定的个人信息。用户也可以之后通过撤回许可来注销这些小程序。不同小程序对用户来说，在必要性和要求用户提供数据的敏感程度上差别很大，而这些正是用户可以选择的。将这些差异与用户特征和选择组合起来，我们进行了迄今为止最大规模的，关于消费者隐私决策的大数据研究。

当用户有权选择是否分享个人信息，从而获得小程序服务时，他们会做出怎样的选择？结果显示，如同其他国家一样，中国用户普遍在意自己的隐私。但当面临选择时，绝大部分用户会选择分享个人信息，以获得服务带来的福利。具体而言，当面对数据要求时，75% 的情况下用户会选择给小程序授权信息，并且后续的退出率较低（每月 0.12% 的用户选择退出对小程序的个人信息授权），且随时间进一步降低，显示出他们大多数并不后悔自己的选择；这些比例和欧美用户的行为规律相当一致。另外，用户会在信息敏感度和服务质量之间做出取舍，面对隐私事件时，他们会倾向于用脚投票，提高退出率。随着用户经验的不断积累，他们一开始会更谨慎地选择，但日积月累，他们处理分

享个人数据的经验更多，拥抱的数字服务也会更多。这些行为模式适用于不同的性别、年龄和教育程度。

这些结果表明，**“隐私悖论”的本质是，与个人数据相关的消费者权益具有双重性，一个是隐私被保护的权益，一个是因为分享数据而获得（更好）服务的权益**。两者之间的权衡，才是对消费者权益的完整理解。研究表明，最担心隐私的用户，恰恰是使用数字服务更多的用户（Chen et al.，2020）。所以解决“隐私悖论”的方法，不是就隐私谈隐私，把数据锁起来，而是在保护好隐私的基础上，鼓励数据的流动，这样才能真正让消费者受益。

1.2.2 数据的价值来自何处？

大多数用户愿意与服务提供商共享个人信息，从而享受数字服务的好处。这自然而然地引出了我们在第三章中讨论的问题：数据到底给用户带来什么价值？为什么用户愿意分享数据？

我们总结了在线数据共享的价值，至少表现在三个方面：连接、决策和信任。首先，如我们在《新普惠经济：数字技术如何推动普惠性增长》（罗汉堂，2019）中提到的，**数据分享会增强连接性**。在数字技术的帮助下，数据的产生和分享是如此便捷，普惠性连接达到了前所未有的水平，重新定义了市场以及人们组织生产和协作的方式。一个例证是，由于在线市场的出现，交易的范围、深度和广度都得到了极大提升。传统线下交易的特征一般可以用经济学中的“重力模型”来描述，即本地商户的用户绝大多数来自方圆 10 公里范围内，距离越远，交易越少。而中国当前电商平台上呈现的景象则完全不同。除了生鲜食品，买家和卖家之间成交的平均距离接近 1000 公里，超越传统线下市场服务范围两个数量级，“重力模型”被彻底打破。从连接买卖双方的情况看，10 亿淘宝用户中，月度活跃买家超过 7 亿，同时有超过 1000 万家初创企业和公司作为卖家，其中约一半的创业者是女性。在产品丰富度方面，消费者在线上可购买 10 亿种以上的商品和服务。这个市场的形成，是以参与各方愿意分享数据为基础的。**用 North 和科斯的话说，如果没有数据分享，就没有可以协同的经济活动**。

一个有趣的问题是：如果没有基于个人数据的用户特征，商户不懂用户，会发生什么？具体而言，在今天的线上购物环境中，因为用户面对的是上亿件商品，没有推荐很难找到心仪的产品。如果剔除根据用户个人数据产生的推荐信息流时，会对在线市场产生什么影响？我们尝试通过一个大规模随机试验来回答这一问题。在实验中，我们关闭了用户的个人数据算法推荐系统。结果发现，个人信息的缺失会对买家和卖家产生巨大冲击。由于缺乏个人数据，个性化服务无从谈起，平台推荐只能盲目地集中到那些交易量在前 1% 的品牌所提供的产品上，回到了传统线下市场的营销推广模式。实验结果显示，用户满意度不够导致交易量暴跌 86%，尤其对知名度低、销售额少的小微企业不利。由此可以得出一个重要结论，匹配用户数据与产品，可以大大降低搜索成本，尤其是当市场存在海量产品的时候。当个人数据这一重要的信息源被切断时，消费者在选择潜在商品时只能依靠传统的供给侧的信息源：品牌、信誉和商品一般特征。因为这些来自传

统渠道信息的有效性不足，市场规模大幅萎缩。这一结论与搜索领域的学术研究不谋而合。大量论文证明，即使较小的搜索或匹配成本也会导致商品和劳动力市场的厚度和广度产生剧烈变化（Stigler，1961，1962; Diamond，1971; Pissarides，2009）。

第二，数据分享可以改善决策。海量多种类数据相连接，可以帮助无数消费者和生产者做出更明智的决策，促进更快、更有效、更多的创新产品和服务，数字时代之前不可能出现的商业模式以及新的产业组织形式也随之出现（见第六章对熊彼特式竞争的讨论）。由于无法和大企业一样进行大规模市场调研，中小企业在传统市场中一直难以获得市场和消费者信息。因此通过数据分享改善商业决策对中小企业尤其意义重大。其中一个案例是淘宝和天猫平台上的“生意参谋”，类似服务也可以在国内外平台上看到。这项服务为所有在线商家提供多种信息分析工具，包括卖家自身历史业绩、市场趋势以及潜在消费者喜好等等。大多数生意参谋的新用户是中小企业，它们的销量通常会在开通服务的第一周出现跃升，并在之后的 10 周，已经开通服务的用户和未开通服务用户的业绩差异会逐渐稳步拉大。“大数据”的出现让中小企业获得了以往只有大企业才能享受的先进分析工具，帮助它们快速增长。

在金融领域，数据分享可以改善金融风险甄别能力。传统金融一直难以克服普惠性不足的顽疾，让抵押品不足的低收入人口和小微企业获得融资，而数据分享有望解决这一难题。通过获取借款人的消费和经营数据，已经足够说服贷款人在无抵押的情况下提供金融服务并承担相应风险。通过这种方式，小微企业也可以享受到金融服务。正如诺奖得主 Holmström 所言：“信息已经成为新的抵押品”（Holmström，2018）。大数据让过去无法实现的大规模小微信贷成为可能。2011 年以来，网商银行为超过 2000 万家中小微企业提供了无抵押贷款。网商银行最早推出的“310 模式”已经广为人知，并且现在很多银行都普遍使用：3 分钟申请贷款，1 秒钟能及时到账，0 人工干预。这种由大数据风控支持的小微贷款，为千万计创业者带来了机遇，这也是罗汉堂《新普惠经济：数字技术如何推动普惠性增长》的主要发现之一。

第三，数据分享可以建立信任。新型线上市场有数以亿计的参与者，要像线下市场面对面交易一样顺畅无阻，对产品及参与者的信任机制必不可少（Tadelis，2003）。有了线上的数据分享，消费者就能对商品和生产者进行评价，而生产者则通过这样的评价系统，努力打造良好的信誉。所有参与者都是数据的生产者，也同时从数据的交换中受益。与之形成对比的，是数字时代之前，诺奖得主 Akerlof 描述的“柠檬市场”，即消费者和生产者信息不对称，消费者缺乏对产品的信息和信任，只愿意选择低价产品，从而劣币驱逐良币，赶走了好的服务商，只剩下质量不好的“柠檬”，随之恶性循环，直到整个市场消失。而通过线上评价系统，一方面数据分享让新的卖家获益，另一方面高质量卖方也可以通过重复交易，让自己与那些低质量、“一锤子买卖”的“柠檬”商家区别开来。随着时间积累，这些卖方的品牌脱颖而出，可以获得更好的销量。在这个过程中，所有参与者都是数据分享的受益者。

大数据往往可以用三个 V 来概括：即**数据量**（Volume）、**多样性**（Variety）**和速度**（Velocity）。数据量指的是能观察、记录、处理和分析海量的数据。多样性代表数据的宽

度，即能处理许多不同类型、不同维度的数据，从而满足数字市场中卖方和买方的不同需求。速度是指收集、处理、分析和使用数据的速度在不断加快，也就是实时性。

结合前面的讨论，大数据的两个 V，即大容量和多样性数据正在彻底改变人类的交互和协作。这是因为数据可以改变消费者与生产者之间的连接方式，增强买方和卖方之间的信任，并且让决策变得更迅速、更明智。同样关键的是，这些基于数据的连接、信任和决策过程，正在以前所未有的速度，甚至是实时地进行：与实体商品不同，数据只有流动起来，才能传递信息，创造价值。**大数据的三个 V 向我们展示了数据如何创造价值：海量且多维的数据实时地驱动社会经济活动**。这正是数字经济的本质，而所有的经济活动的参与者都是受益者。正如哈耶克所洞察的，开放且充满竞争的市场，加上来自各方的信息分享和决策，才能让整个社会受益。

1.2.3 如何缓解隐私风险？

尽管数据分享可创造巨大价值，但也存在风险。数据创造的价值越大，保护隐私和数据安全的紧迫性就越高。数据生命周期的每个阶段，从收集到集成，从分析到应用，都存在数据泄露和隐私风险。个人有知晓和拒绝数据收集的权利，这是广为接受的观念，然而在现实中，要防止个人信息过度暴露和信息泄露是一个艰巨的挑战。2017 年全球数据外泄和遭窃取的记录达到 16 亿起，造成巨大的经济损失，引发了消费者对隐私问题的极度担忧。近年来，如 Facebook 和剑桥分析的数据滥用事件引发了社会的广泛关切。

当下社会关注的热点是如何通过法规保护好隐私，而同样需要关注和理解的是行业和企业的隐私保护措施。因为数据分享和运用是经济活动不可分离的一部分，法规只能规定经济活动的边界，弥补市场失灵的部分。只有当行业和企业把个人隐私保护和数据安全作为商业的一个重要条件，并配置相当的机制和技术，才能真正实现目标。

在这个维度上，全球很多行业和企业已经在做大量的探索。我们在第四章总结了企业做好隐私保护的逻辑和实践。**简而言之，有效保护隐私的逻辑，是将隐私工程化（privacy engineering）和隐私增强技术（privacy-enhancing technologies，PETs）结合起来**。隐私保护工程化，是指将个人隐私保护的法规和原则，融入到产品设计中来开发和使用软件应用。例如在用户交互设计上，隐私工程可以加强用户对隐私条款的理解，增强对敏感信息的控制。

隐私工程化可以应用到大数据生命周期的各个阶段。在信息收集阶段，企业必须获得用户的许可，并且必须遵循收集数据的必要性原则。在集成和存储阶段，企业处理数据前要可以过滤敏感信息。这些信息还可以加密，这样即便出现数据泄露的情况，个人信息也不会被滥用。脱敏和加密后的数据，可以用于分析，了解消费者及其需求，并且在严密且持续的风险管理之下进行。最后，要可持续且高效地使用数据，企业要在隐私保护需求和用户数据许可最小化之间取得合理的平衡，这样才能既保护隐私，也不至于因为过分许可打搅用户。最后，用户还应该保有个人信息的删除权。

可以看到，隐私工程化意味着需要很多隐私保护技术，从而防止不可信或潜在恶意的数据收集者侵害用户的隐私。例如多方计算技术（MPC）可以让数据分析者从数据中

提取有用洞察，却不会暴露或回溯至原始数据。区块链技术也可以通过对个人数据进行加密和密钥，降低隐私风险。这些技术的目的，是让服务提供方进行大数据分析时，懂得客户特性和需求，然而却“不知道客户是谁”“数据可用不可见”，从而更好地满足客户和数据相关的两个权益。另一方面，隐私工程和隐私技术成本不菲，给初创公司和中小企业带来更多挑战。在多大程度上做好，能够同时满足消费者和生产者的需求，从而发挥数据作为生产要素的价值，同样是值得整个社会讨论的问题。

长期来说，食品安全以及飞行服务等行业的历史经验表明，假以时日，合理的机制设计和不断完善的技术，可以在很大程度上缓解数据隐私和安全问题。就像吃很多东西不见得会中毒，频繁乘坐飞机不见得会出事，数据分享的体量和隐私及数据安全并不是必须的取舍。当技术足够强大，机制足够合理，今天看起来严重的隐私风险即便无法彻底杜绝，也可以得到有效控制。

1.2.4 如何从数据的本质看数据的权属、利益分配和保护责任问题？

我们在报告第五章中提出了一个数字时代理解数据和隐私本质问题的综合框架，也称为“数据权衡框架”。数据的问题需要综合视角去理解，否则很容易陷入“盲人摸象”的困局。首先，数据具有和物理商品截然不同的本质属性，在生产和使用过程中牵涉到多方。其次，我们需要综合考虑用户和数据相关的两个福利，即隐私保护权益和因为分享数据而获得服务的权益。再次，数据分享在经济活动和人类协同中必不可少，数据只有通过在社会和经济行为中流动才能创造价值。

这个数据权衡框架包含数据的两个本质特征、三个视角，以及一个基本原则。

1. **数据的 2 个本质特征：非竞争性和不可分离性**。首先，数据和物理商品不同，据有非竞争性，可以被无数次生产和使用，而不会消耗数据相关的对象。其次，不管数据的使用者是谁，都可能对数据相关的主体带来潜在影响；数据使用和数据主体存在不可分离性。

2. **数据生产和使用的三个视角：数据生产者、数据主体和使用场景**。这里数据生产者（在商业环境下）是指观察、收集和处理数据的机构和个体。数据主体是指数据描述的个体（个人数据）或对象（非个人数据）。使用场景是指使用数据的经济或社会活动。

数据需要被观察才能产生，所以数据相关主体并不一定是数据的生产者。基于数据的两个本质特征，数据生产者和数据主体的利益是相互关联的。

从数据生产者的角度来说，要产生数据，既需要数据主体，也离不开数据生产者，并且数据的使用也会同时影响两者。数据的非竞争性本质决定了，数据可以有无数个所有者，而不会消耗数据或者数据主体本身。举例来说，一个人发表演讲的数据是由每一个听众分别产生的，并且会因为每个听众的视力、听力和关注点不同而有所差异。基于数据形成的信息也可以分享给不在场的人，而不会损耗演讲者。

从数据主体的角度出发，使用数据会对他们造成影响，因此他们的权益必须得到保护。

从使用场景的角度看，数据不应被简单地类比为一种有固定价值的商品。在实践中，

数据需要归集、存储、分析，形成对场景需求有价值的信息洞见。一方面，这个过程需要消耗成本和能力。另一方面，数据的价值是变动的，取决于基于数据的信息能在多大程度上提高经济和社会活动中交互的效率。所以，数据的使用是经济活动中不可分离的部分，其价值也随着具体的场景需求而变化。

数据的两个本质特征，和数据生产者、数据相关主体、使用场景三个视角，可以帮助我们理解数据和其他商品或生产要素有什么不同，数据是如何产生、如何发挥价值的，以及牵涉到的相关方。基于这个权衡框架，我们可以得出三个结论：**首先，将数据等同于一般商品那样拥有唯一所有权的观点是不合理的。第二，隐私保护的重点应该是在数据使用中尊重和保护数据主体的隐私权，而不是将数据独有权给予数据相关主体，否则难以发挥数据作为生产要素的价值，最终让所有的相关方受损。最后，数据生产者和数据相关主体之间应在平等、自愿的基础上达成协议，从而双方都可以从数据的生产和使用中受益。**

总结起来，数据治理的核心原则应该是在促进数据流动的同时保护数据主体的权利。

1.3. 数据治理问题

1.3.1 数据治理的演进逻辑

我们在报告的第六章讨论数据治理的几个核心问题。

我们的数据权衡框架有助于更好地理解数据隐私监管和治理的演进。上世纪 70 年代，美国颁布了《公平信息实践》（FIPS），为现代隐私保护的治理法规和监管奠定了基础。《实践》基于五大原则包括：（1）通知/知情；（2）选择/许可；（3）接入/参与；（4）完整/安全；以及（5）执行/纠正（美国联邦交易委员会，1998）。

这 5 大原则指导美国联邦委员会“鼓励和推动有效的自我监管，作为保护消费者线上隐私的主要方式”（FTC，1998），进而成为后来隐私和数据治理条文和法规的原型，包括欧盟的《数据保护原则指令》（DPPD）、《通用数据保护条例》（GDPR）以及美国的《加利福尼亚消费者隐私法案》（CCPA）。**以 FIPs 为基础演化而来的个人数据治理法规有一个关键的共识，那就是不要将数据锁入“保险箱”里或拘泥于数据所有权，而是鼓励安全的数据流动，同时保护消费者隐私。**例如，经合组织（OECD）就提出“兼顾平衡的隐私立法，并保护相关人权……同时避免干扰……数据的流动”。让消费者从分享个人数据中受益，这个定位和我们的数据权衡框架中确立的基本原则是一致的。

虽然在大原则上一致，但在数据使用上有不同程度的限制，反映出不同国家和机构对数据治理的实践差异。诺奖得主 Elinor Ostrom 指出，尽管动机良好，但过于严厉的治理政策会对专利、知识产权、授权、定价甚至“数字经济的存续”带来不利影响。她认为，政策制定应让各参与方都成为利益相关方，从数据分享中受益（Layton，2019）。很多其他学者也得出了相似的发现和结论（Goldfarb and Tucker，2011；Martin and Murphy，2016）。“任何条例都会带来应用的成本，数据隐私法可能会限制宝贵的信

息流动，带来隐私和安全风险，提高市场进入障碍，增加创业者的不确定性，以及催生寻租行为”（Layton，2019）。

1.3.2 数据和竞争的关系

数据驱动的商业行为在竞争中会扮演越来越重要的角色，因此我们需要理解数据驱动的市场行为会如何影响竞争。竞争法的核心目标是：“确保消费者可以从竞争的力量中受益”（Shapiro，2018）。要判断竞争是否被扭曲，消费者的利益是否受损，我们要用事实说话，深入理解行业结构、企业的商业行为，并评估其市场表现。

我们先回顾一下交易对竞争以及国家的企业竞争力的正面影响。如上文所述，在中国，线上市场的出现让买卖双方的平均距离从 10 公里上升至 1000 公里。在 1776 年出版的《国富论》中，亚当 · 斯密指出，垄断力量是“良好管理的大敌”，而交易范围的扩展会打破这种垄断：

“状况良好的公路、运河、船只往来的河流，这些大大降低了货运的开支，将一国偏远的地方与城镇周边置于同一发展水平上。它们都得到了最好的交通改善。这将刺激边远地区——一国最偏远一环的发展。通过打破本地商人的垄断，对城镇的发展多有裨益，也对国家的其他地区有益。尽管交易扩张在旧市场中引入竞争性的商品，但也为旧产品带来了新的市场。此外，垄断是良好管理的大敌。只有在自由和全面的竞争下，市场中的每个人为了自我防御都要遵循规则，好的管理制度才能建立起来。”

如同亚当 · 斯密时代的“公路和运河”延伸了贸易距离，打破了垄断，提升了偏远地区商家的竞争，21 世纪出现的平台数字网络也打破了本地垄断力量，并且如下文所述，建立了一个更健全、更具竞争性的商业环境，只不过两者的速度和效率无法同日而语。

另一方面，数字经济中可能妨碍竞争的商业行为在全球范围内正在引起越来越多的关注和争议。由于篇幅所限，本报告无法为所有相关问题提供答案，我们基于初步证据，聚焦于讨论三个和数据竞争相关的关键问题。

首先，企业在多大程度上利用大数据技术歧视性对待消费者？

商家今天获得关于客户的信息前所未有，无论数据量还是数据种类都远超前人想象。从理论上讲，企业通过数据更懂消费者，是有可能成体系地对消费者采取价格歧视，攫取消费者利益。在实践中，我们也会在不同国家看到“大数据杀熟”的个别案例。但总体而言，和蓬勃发展的大数据运用相比，大数据杀熟的案例要少得多。也没有任何国家的研究发现大数据杀熟的广泛现象，反而免费、普惠成为更加普遍的时代特征。这是为什么呢?

一个可能的解释是，数字技术改变了生产者和消费者的关系。其中一个结果，随着累积这些高颗粒度的数据，商家已经不再追求从单独的产品或服务中获取最大利润，而是追求提供一个以客户需求为中心的综合服务，建立更高的客户忠诚度。普惠性就是尽可能扩大客户的多样性和数量，今天已经成为越来越多企业核心的商业目标。例如 Ichihashi 在 2020 年的研究显示，很多数字平台希望向平台上的卖方公开买方的特征信

息，而不是将买家隔离开，进行经济学家所谓的“价格歧视”——以不同的价格向不同的群体出售类似的产品，以便从支付意愿最高的消费者身上获取最大的利润。实际上，要成功地进行价格歧视，卖方需要有能力将不同群体的消费者隔绝开。而今天的互联网让信息隔离越来越难，消费者可以在非常广阔的范围内搜索不同的卖家和价格，让竞争越来越激烈。

价格歧视并非唯一的顾虑。美国的消费者保护机构，例如联邦贸易委员会担忧“企业可以使用大数据，让那些低收入、缺乏服务的群体失去获得信用服务和就业的机会”，并表明要对这种不公平的行为进行指控。但现实中数字金融不断增强的普惠特性，将缓解这样的担忧。至少在肯尼亚、中国和其他很多国家，普惠金融取得了突飞猛进的发展。所以，在我们关注大数据杀熟的案例的同时，我们也不应该忘记这样的案例并非多数，也要具体分析其背后的原因。

其次，大数据是否在妨碍竞争，让市场出现赢者通吃的局面？

关于数据竞争的一个普遍担忧是，网络平台的外部性（直接和间接的）和规模经济可能带来市场进入障碍，从而引发赢者通吃的结果。从具体效果看，美国科技产业的一些领域的确出现了市场集中度提升的现象，但中国却上演了一个非常不同的故事。在中国，数据驱动的产业出现了强竞争的特征，潜在的新竞争对手常常让在位的互联网企业夜不能寐。

例如电商行业，尽管在线零售占社会总零售比例已经超过 27%，但由于激烈的竞争，阿里巴巴这样的具备先发优势的领先者，面对市场新进入者和原有竞争对手的快速增长。在短短四年内，阿里巴巴在中国电子商务销售额占比从 2015 年的 78% 降至 2019 年的 56%[1]。拼多多作为市场新进入者，在 3 年内吸引了超过 4 亿用户，销售额增长超过 100 倍。

中国曾经最典型的大数据企业百度，长期统治中国的搜索引擎市场，并且曾在大数据和人工智能领域遥遥领先，2010 年它的市值大于阿里巴巴和腾讯，然而如今却远远落后于这两家公司和很多新进入者。抖音的母公司字节跳动异军突起，只用 7 年就超越百度成为互联网广告收入领头羊。京东商城占中国电子商务销售额 17%，它的投资者包括谷歌和沃尔玛这样的巨头，最近成功“占据家用电器市场全渠道最大市场份额”[2]。在移动支付市场，类似的激烈竞争也在频繁上演。这些证据表明，我们很难得出先发的大数据优势会引发“赢者通吃”的结论。

一个可能的解释是，大数据只是商业竞争的一个要素，并不必然决定行业的格局和企业的命运。首先，在数字经济中，数据的使用只是商业模式运行的一部分。尽管在今天的商业模式中，数据扮演了比以前更加重要的角色，但企业间的竞争仍旧由商业模式及其执行决定。另外，与传统经济不同，数字服务的消费者可以选择多个平台——用户

[1] 参见文章Alibaba，JD.com Lead in China，but a Few Others Are Making Dents，too 和Retail and Ecommerce Sales in China 2018。

[2] 参见 Wikipedia 词条"List of largest Internet companies"，文章The biggest ecommerce companies in China —a brief guide和JD.com's Market Share Tops in All Channels of the Home Appliance Market。

可以选择提供类似服务的不同供应商，从而多方分享自己的个人数据。再次，随着数据的增加，数据的边际效用会随之下降。研究表明，如果一种资源能为企业带来垄断的竞争优势，那么这种资源一定是无法模仿的、稀有的、高价值的且可持续产出的，“然而通常数据并不具备上述任何一种特质”（Lambrecht and Tucker ，2017）。

显然，在数字经济的发展过程中，存在妨碍竞争的企业行为，需要通过法规纠正，这也日益成为各国社会关注的一个重点。但这些现象，和所谓的“赢家通吃”，无论在理论和实践中都没有必然的联系。我们对事情性质的判断，应该基于事实，基于不同国家、不同行业的实际情况，而非基于假设的“黑板经济学”，否则会适得其反，妨碍数字经济的发展。

再次，大数据在多大程度上在给创新带来障碍？

从事实上看，数字技术在已经带来深度影响的行业，包括媒体、社交媒体、电子商务、金融、视频、出行和共享单车等等，几乎每一个行业都有新的商业模式和新的市场进入者。在数字技术和数据的帮助下，这些新的商业模式为行业带来了“破坏性创新”，也逼迫那些墨守成规的成熟企业做出改变。在全球几乎所有市场，带来激进创新的企业都是那些资金和资源非常有限，但拥有无限想象力和雄心壮志的初创企业。可以说，创新已深深烙印在科技初创企业的 DNA 中。

尽管有合理的疑问，但我们认为有几个原因可以说明，为什么大数据天然会推动产品和生产流程的创新。

第一，大数据的三个 V 已经成为重要的生产模式和商业模式的创新引擎。企业连接和了解客户的能力大大加强，因此可进行更明智的决策，并进行创新试验。这是在数字技术深度改变的行业中，创新层出不穷的重要原因。

第二，大数据的三个 V 带来了前所未见的大规模、深度合作，这也大大加速了创新在市场中的部署和应用。平台作为连接供给和需求的载体，成为创新扩散的重要推动力量。平台处于竞争压力，会愿意改进商业基础设施，推动平台上企业的创新发展。实际上平台上的创新非常活跃，例如新品牌和小众品牌在各类平台上的爆发式增长，就是显著证据。

本报告的结构安排如下，第一章总结本报告的核心内容。按照逻辑递进的方式，第二章基于消费者在隐私相关决策时的实证分析，推导出不要单维度看待隐私，而需要综合理解消费者和个人数据相关的权益。第三章谈论数据分享的价值源自何处。第四章介绍从行业和企业的角度，做好隐私保护的核心逻辑和方法。第五章提供了一个基于数据本质特征的综合理解数据权衡的框架。第六章运用这个框架理解全球数据治理的相关问题，并且就数据和价格歧视、竞争、创新的关系，基于实证研究进行了讨论。第七章提供了简短的总结思考。

第 2 章 从消费者权益的视角理解隐私悖论

毋庸置疑，隐私保护对数字经济的发展，尤其是大数据技术的应用至关重要。关键问题是：对消费者而言，何种程度的隐私保护是足够的？消费者在真实生活中，是如何决定是否分享个人数据的？隐私保护和个人信息分享之间是否存在权衡取舍？

2.1. 让人费解的隐私悖论

消费者分享个人数据的意愿究竟有多高？他们在多大程度上在意隐私保护？令人费解的是，人们对这些问题的回答，与他们的实际行为往往相互矛盾。学术界称这一现象为“隐私悖论”：大多数人表示自己在意隐私问题，却常常免费或在很少的经济补偿下，分享自己的信息（图1）。

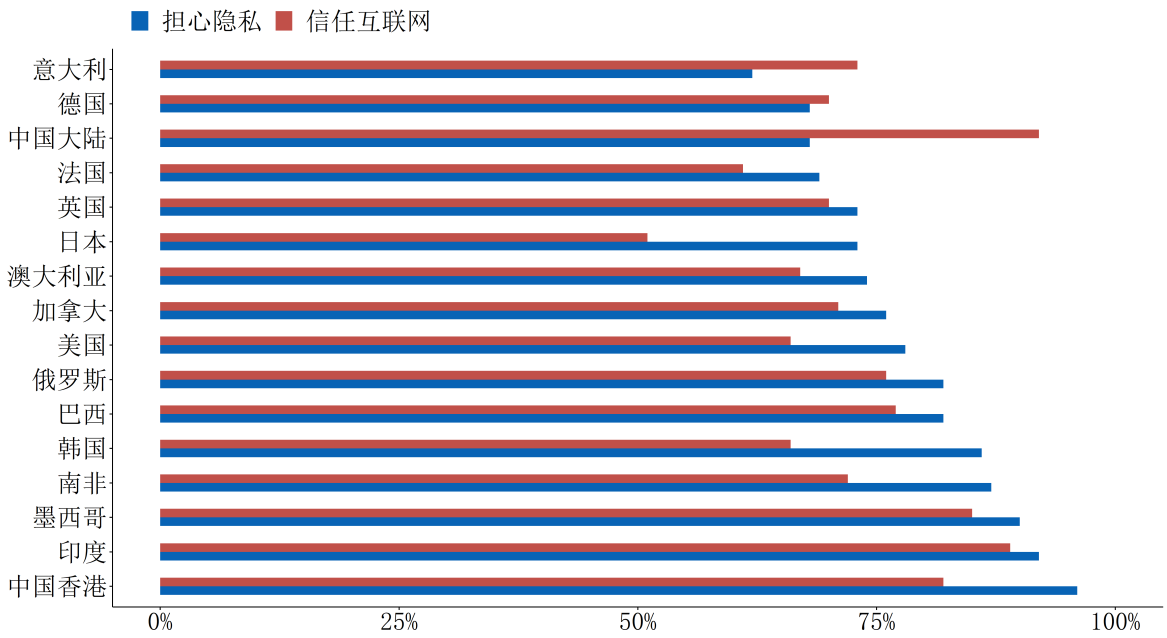


图 1: 担忧很多，拒绝很少

资料来源：CIGI-IPSOS，2019，全球互联网安全和信任调查 Global Survey on Internet Security and Trust。

注：该图表显示了全球互联网用户中担忧隐私泄露风险的比例（蓝色）与信任互联网从而愿意分享个人数据的比例（红色）。

这种言行不一的行为模式普遍存在于全球各国，不分国家和文化。Eurobarometer 一项关于数据保护的调查显示，欧盟地区 67% 的公众担忧自己在互联网上无法控制个

人信息[1]。美国皮尤研究中心的一项调查发现，美国 93% 的成年人认为有必要了解谁获取了他们的个人信息[2]。中国消费者协会（2018）指出，80% 的调查对象认为他们的数据发生过泄露，并且被广告和销售电话骚扰。Global Privacy Enforcement Network 在 2018 年的一项调查显示，只有 28% 的人对移动、宽带和通信设备运营商的信任度较高，对社交平台的信任度更低，只有 15%。

尽管对隐私保护的焦虑广泛存在，但很多人还是愿意分享自己的隐私数据，往往是在没有或很少的经济补偿的情况下。当分享他人信息时，人们几乎无所顾忌。例如，Athey 等（2017）研究者发现，因为一小块比萨，大多数参加他们实验的人就愿意提供自己朋友的电邮地址。

隐私悖论让学者和政策制定者困惑不已，难以理解为什么全球数十亿用户在分享个人数据时会如此轻率，尤其很多人都曾在调研中表达过对隐私的忧虑。我们可以理解用户很难拒绝使用诸如 Facebook 这样的主流应用，但互联网上每天都会产生大量新的应用，用户也往往 “来者不拒”。隐私悖论体现了用户在对待隐私问题上，言行之间的巨大差别。正确地理解隐私悖论，对理解消费者对个人数据分享涉及的权益诉求至关重要，是制定隐私保护政策所需要考虑的关键因素。

2.2. 通过大数据研究揭示用户的隐私决策

目前对隐私悖论有几种不同的解释。一种观点认为，这是因为当事人不了解侵犯隐私可能带来的严重后果。还有一种认为，由于目前主流隐私条款只能 “接受或退出”，一旦缺乏其他可替代服务，用户不得不让渡一定的隐私权（Chen and Michael，2012）。随着市场竞争越来越激烈，很多可替代的产品正在涌现，这种解释越来越缺乏事实基础。另外，认为用户不了解分享的严重后果和用户对隐私的普遍担忧是相悖的。第三种观点则认为，当面临真实选择时，是人们的行为，而不是口头表达，揭示了消费者对待数据分享的真实态度。

因此，理解隐私悖论关键在于，当有选择自由时，用户在真实环境中行为如何。利用支付宝数据，我们进行了一次大规模实验。支付宝目前的活跃用户超过 10 亿，并且在平台上有几百万个小程序，有的来自小型初创公司，有的来自成熟大企业。为了使用这些小程序，用户必须同意分享自己的个人数据。授权同意之后，用户也可以通过撤销授权来退出。不同小程序要求的个人数据，在必要性和敏感度上各不相同。将这些差异与用户特征和选择相结合，**我们进行了一项迄今为止规模最大的、关于隐私决策的大数据研究**。

具体来说，我们聚焦小程序用户的授权和退出行为。研究小程序的授权和退出有几个优势。首先，它反映了用户面临个人数据分享问题时的真实决策，而不是基于问卷调

[1] 参见 2015 年欧盟报道，Agreement on Commission's EU data protection reform will boost Digital Single Market。

[2] 参见皮尤研究中心报道，Americans' Attitudes About Privacy, Security and Surveillance。

研。基于问卷的研究，有一个明显的弊端，就是很难让调研对象对所提问题感同身受，做出真实的选择。例如，汽车购买者常常在调研中夸大对燃油经济性和小排量发动机的偏好（Anwyl，2011），这是因为在调研中，消费者不用面对真实的成本，包括小排量发动机带来的空间、舒适性和安全性上的损失等。

第二，支付宝用户涵盖了不同性别、年龄、收入和教育背景的群体，能比较全面地反映不同用户的行为差异。

第三，我们的研究覆盖超过 5 万款的小程序，它们提供的服务和要求分享的个人数据范围有很大不同。尽管用户很难绕开微信支付或支付宝，但对于是否使用这些小程序，用户有足够的自主选择权，例如，他们可以决定是否为享受餐厅的订餐服务而分享自己的个人信息。

案例 1 显示，在支付宝平台使用特定服务时，不同用户在决定分享个人信息时有不同的行为模式。这些行为模式与罗汉堂撰写的另一篇论文的结论一致（Chen et al.，2020）。

案例 1 支付宝小程序

支付宝平台的活跃用户超过 10 亿人，并且拥有几百万款小程序。这些小程序提供的服务几乎无所不包，从外卖到金融服务应有尽有。

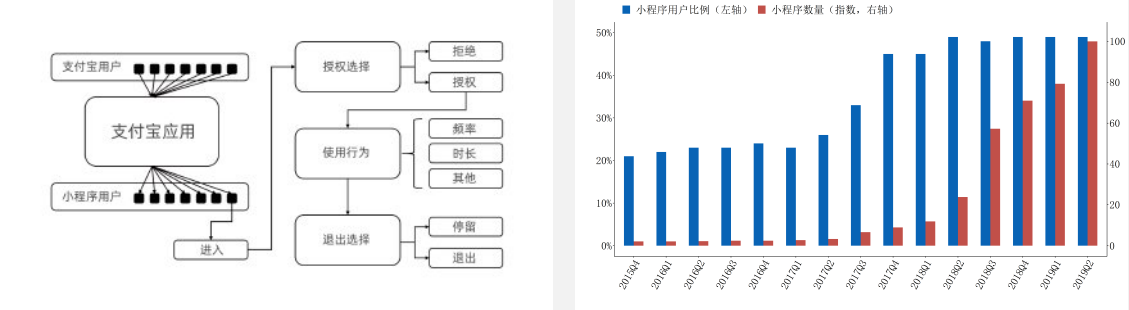


图 2: 支付宝小程序

不同小程序需要的信息内容和信息敏感度各不相同。一些小程序只需要提供用户姓名、位置和联系人列表等，还有一些则要求提供用户的信用分。结合用户不同的特征，这些差异可以提供丰富的洞察，了解用户在小程序提供的福利以及隐私数据之间如何进行权衡选择。

小程序提供的服务涵盖交通、娱乐、政务服务、餐饮和金融等等。例如，共享单车小程序获得用户的信用数据后，可以让他们免押金租用单车。近年来，支付宝上小程序数量快速增长。从 2015 年到 2019 年，小程序用户渗透率从 21% 上升至 49%，用户规模上亿计。

在使用前，小程序会要求用户进行数据授权。这些数据包括用户姓名、昵称、性别、手机号、地址、网络信息、芝麻信用、支付账户、转账信息和相关服务信息

等等。用户可以自由地进行授权和退出。

当用户有选择权时，他们是否愿意分享个人信息，从而获得小程序提供的服务？

我们发现的第一个行为模式是，当面临选择时，用户很乐意授权个人信息。如图3所示，不同类型消费者小程序的授权率在 64%-86%。平均授权率超过了 75%。男性、教育水平高和年轻用户分享个人信息的整体意愿更高。在不同性别和教育水平的群组里，授权率差异较小。拥有本科（以上）学历的用户，接受小程序的意愿更高。我们通常猜测，教育程度低（对隐私泄露严重后果的了解相对更少）的用户更易分享信息，但结果恰恰相反。不同年龄段用户之间的差异最大，45 岁以下用户授权率比 45 岁以上用户高出近 20 个百分点，可能是因为年轻用户对在线服务更熟悉且使用更加熟练。

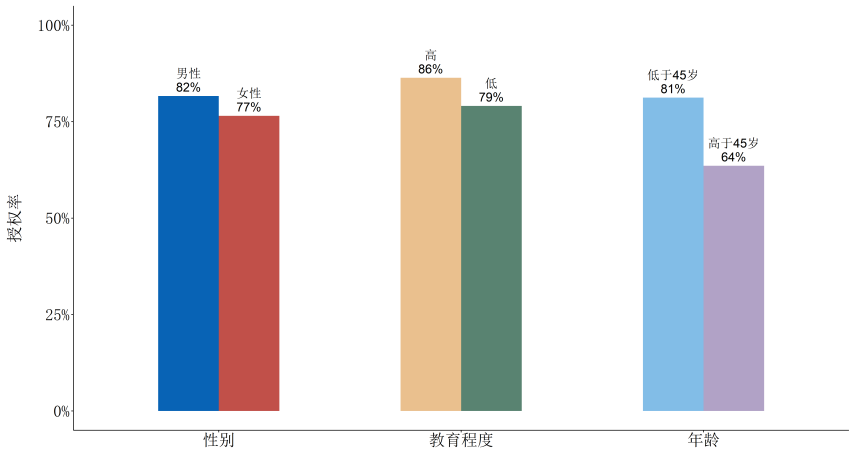


图 3: 谁的个人信息授权率更高?

资料来源：罗汉堂。

注：授权率是小程序的访问者中选择允许小程序获取用户信息的比例。我们将具有博士学位、硕士或学士学位的用户定义为高等教育用户。其余用户定义为低学历。年龄方面，我们以 45 岁为界分成两组。

人们不仅在大多数情况下愿意分享个人信息，使用小程序，而且他们后续也很少退出，表明他们不会后悔自己的选择，至少没有足够的退出倾向。数据显示，小程序的整体退出率非常低，2016 年-2019 年每月只有 0.12% 的退出率。这意味着，大多数用户愿意用个人信息换取服务。一旦做出授权，他们很少改变自己的决定，可能认为退出不太重要，或对获得的服务感到满意。2020 年的一篇学术论文得出了类似的结论，研究者发现，在一个名为 AdChoices 的精准在线广告推送服务中，2010 一年美国用户的退出率只有 0.23%（Johnson et al.，2020）。加拿大和欧盟地区的用户退出率也很低，分别为 0.16% 和 0.26%。

尽管用户愿意分享个人数据，但不代表他们不在意隐私。信息的敏感度越高，愿意分享的用户就越少。我们的研究结果与此前的研究发现相似（Goldfarb and Tucker，2019），用户的隐私忧虑是根据不同因素相应变化的：信息的敏感度不同，用户的隐私

忧虑也不尽相同（图4）。与一般公共信息如昵称、头像相比，对于更加敏感的信息，例如支付宝账号、机动车注册信息等，授权率平均降低了 20%。对于更年长的用户，例如 55-65 周岁，这一差距进一步拉大到 30%。整体而言，人们更在意敏感信息的分享。

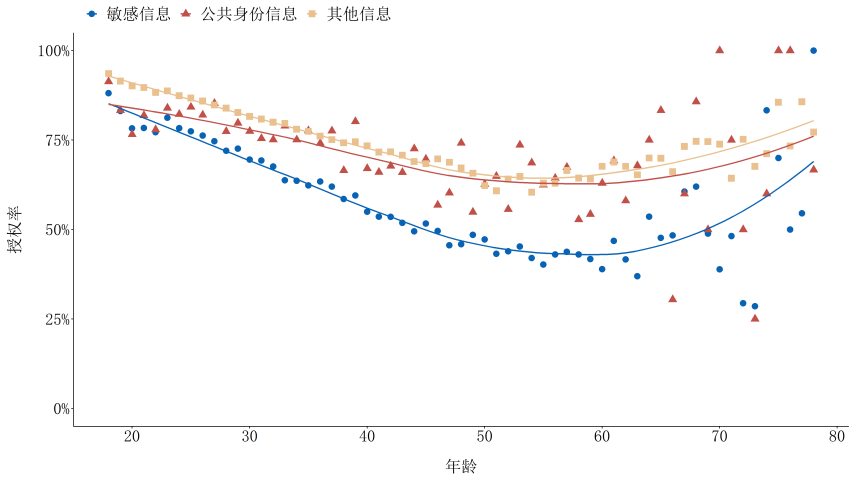
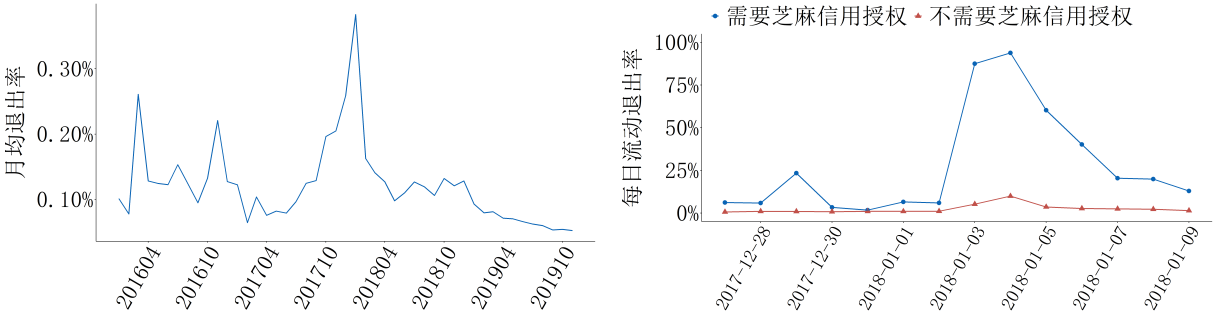


图 4: 数据敏感程度与授权率

资料来源：罗汉堂。



(a) 总退出率　　(b) 敏感数据和非敏感数据退出率对比

图 5: 用户退出率的变化

资料来源：罗汉堂。

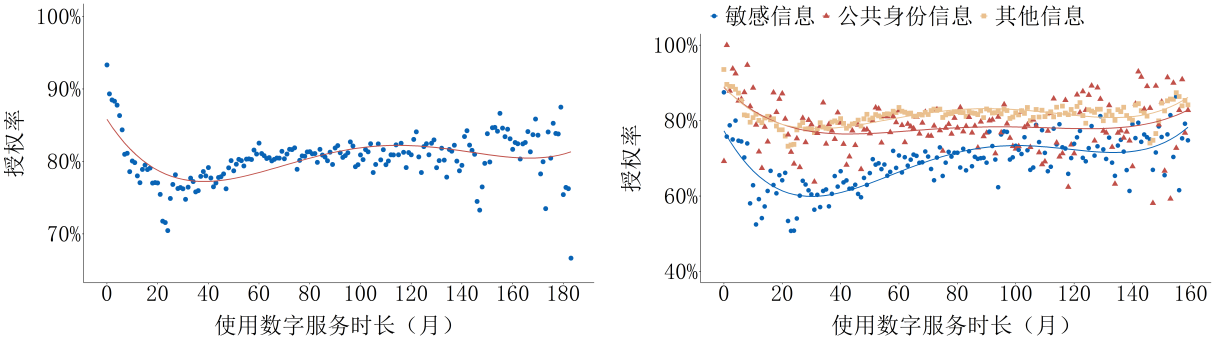
注：（a）总退出率是指月均退出率，即月度退出数除以月度授权数。当用户初次允许小程序访问其个人资料，即视为授权发生。当用户拒绝或取消授权，则视为退出。（b）敏感和非敏感数据的退出率是每日退出率，每天退出人数占每日退出与授权人数总和的比例。

值得注意的是，负面的隐私事件会大幅提升用户退出率。2018 年，支付宝的年度账单事件就是一个例证[3] 。每到年末，支付宝会向用户提供年度账单，来总结他们在一年中的消费。一般情况下，用户乐于与朋友分享自己的消费体验。然而在 2018 年 1 月，用

3 参见《南华早报》报道，Alibaba’ s payments affiliate apologises for opting in users for credit scoring system。

户在支付宝打开 2017 年年度报告时，发现一个新的条款，要求用户授权信息开通信用分。然而这一授权要求在声明中并不足够明显，用户认为应该更加明确地进行通知。此事件后，用户退出率明显提升了，月退出率从 0.12% 上升至 0.3%，尽管之后又迅速恢复至正常水平，并在 2018 年下半年进一步降低（图5）。由此可以看出，尽管用户很少退出，但当他们认为隐私信息授权变更的知情权被损害时，就会“用脚投票”。

随着用户使用数字产品的经验更加丰富，他们也更愿意分享个人信息。此前的研究发现，随着人们对新产品和科技的了解增多，他们对隐私也更加关注（Acquisti et al., 2015；Goldfarb and Tucker，2012）。我们的研究提供了更丰富的理解。我们发现，随着支付宝使用时间的增多，授权率一开始会出现下降，证明了与新用户相比，经验丰富的用户的确会更谨慎地分享个人信息（图6a）。当使用时间超过 40 个月后，授权率会回升至和新用户相当的水平。这种模式揭示出一种 U 形的学习曲线，用户在长期内会接受更多的数字应用。并且，这个趋势对敏感度不同的数据都存在（图6b）。所以，在长期来说，更丰富的数字经验会让用户更加拥抱数字技术，分享个人数据的意愿也更强。



(a) 授权率和使用数字服务时长　　(b) 授权率和所需要的信息

图 6: 数字经验与授权率

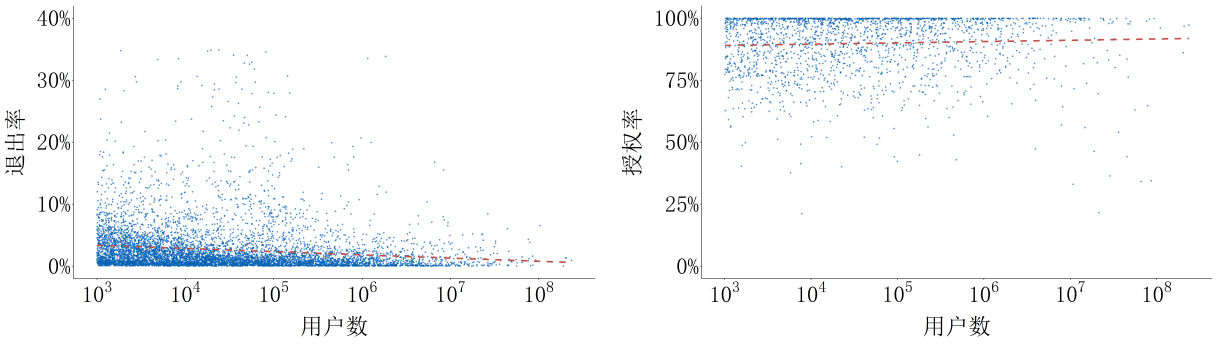
资料来源：罗汉堂。

注：使用数字服务时长指的是距离用户初次注册支付宝的月数。

用户在使用新的小程序时，并不会考虑这些应用的流行度；但那些使用人数少的小程序，用户后期退出的几率会更高（图7）。我们用现有用户数来衡量小程序的受欢迎度。可以看到，用户数和授权率之间并不存在明显的相关关系，这说明当用户在尝试新的服务时，并不会在意使用人数多寡，但用户数少的小程序退出率更高。一个可能的解释是，用户数少的小程序所提供的服务较差，因此退出率更高。

最后，用户对数字平台的信任，比如此次研究中的支付宝，会鼓励他们更多地分享个人数据。2019 年 10 月支付宝小程序进行了一项实验，将小程序用户界面中支付宝标志移除，结果授权率降低了约 3%（图8）。尽管移除标志并没有改变用户和小程序之间的服务条款，却降低了一些用户对小程序的信任。

一个相似的案例是在谷歌的开源安卓市场上，不同主体可以在较少限制下收集、传输甚至交易数据，而用户一般对个人数据的控制较弱。一项研究分析了该市场中 30 万款



(a) 退出率与受欢迎程度　　(b) 授权率与受欢迎程度

图 7: 退出率、授权率、及应用受欢迎程度

资料来源：罗汉堂。

注：我们将受欢迎程度定义为每个小程序用户的对数。

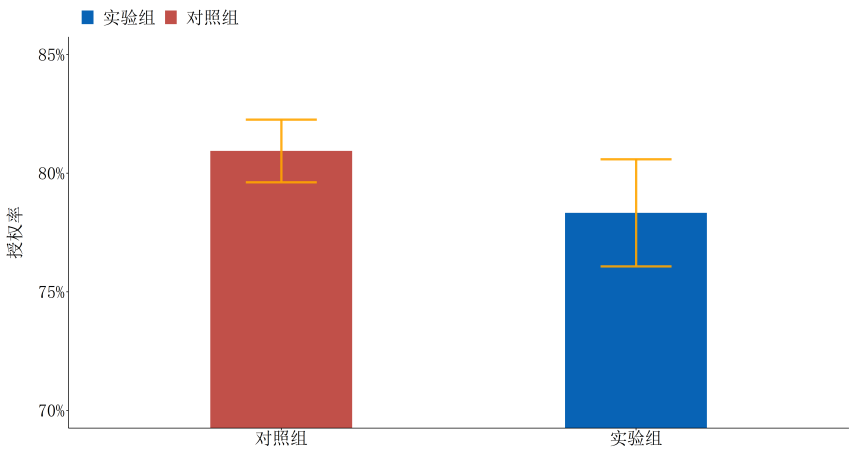


图 8: 实验前后日均授权率变化

资料来源：罗汉堂。

手机应用，发现应用程序开发者的信誉会影响用户对隐私的忧虑（Kummer and Schulte，2019）。

综上所述，对用户隐私决策的大数据研究表明，在“真实世界”中，用户的隐私忧虑的确存在，但与个人数据被肆意滥用而用户别无选择的“假想世界”相比，是有巨大差别的。如同全球其他国家一样，中国消费者也普遍关注隐私问题。但当面临选择时，大部分用户会选择分享一定程度的个人信息，来换取数字服务带来的福利。不同信息敏感度和服务质量也会在他们的决策中起到重要作用。此外，他们也关注隐私事件，并会“用脚投票”。另一方面，他们很少改变自己的选择，表现在退出率较低。数字经验的日渐丰富，会让他们更谨慎地分享更多个人信息，但最终会接受更多的数字服务。这些行为模式超越了性别、年龄以及教育背景的差异。

这些证据说明，隐私保护非常重要，但**只是**消费者在决定是否分享个人信息时考虑的因素**之**一。我们需要一个整体分析框架，来更好地理解数据分享带来的福利以及潜在风险。

最后，我们在另外一篇论文中发现，表达越强烈的隐私担忧的人，恰恰是使用小程序最多的用户（Chen et al.，2020**）**。分享的数据越多，越关心隐私安全，但这并不妨碍他们在获取服务的时候分享数据。隐私悖论并非源于无知或非理性行为。它凸显了数据分享的实际价值，以及我们亟须找到更有效，成本更低的方式去保护个人隐私。如果绝大部分用户实际上愿意分享数据来获得服务，只是他们同时也需要更好的隐私保护，那么最好的隐私保护政策不是将个人数据束之高阁，也不是一味提升数据分享的成本，而是更高效地保护隐私和数据。

2.3. 评估个人信息分享的风险

不同的人对分享个人信息的态度可能截然不同。从 20 世纪 70 年代末期到 2004 年间，Westin-Harris 消费者隐私调查调研了个人对信息技术的理解、质疑和担心。通过调研，他们编制了一个消费者隐私指数，将不同个体分为三类：“隐私本源主义者”：即使可以从中获得更好的服务，这些消费者也不愿分享个人信息；“隐私实用主义者”：他们会根据具体情况进行判断，看获得的服务是否值得信息分享；“隐私乐观主义者”：他们对个人数据的收集和使用并无担忧。每一类消费者的行为都有合理解释，因为他们对数据安全的偏好不同，与获得服务之间的权衡也不同（Equifax-Harris Consumer Privacy Survey，1991）。

不同偏好的消费者占比是多少？全球数据驱动营销联盟（GDMA）[4] 对全球 10 个国家的消费者进行的调查发现，51% 的消费者属于实用主义者，26% 属于乐观主义者，只有 23% 是隐私本源主义者。结合这一发现，我们分析认为，在分享个人信息有足够的好处时，大多数用户是愿意分享的（26% 乐观主义 +51% 实用主义），并且由此来获得服务和福利。（见图9）

[4] 全球数据驱动营销联盟 GDMA （2018），Global Data Privacy: What the consumer really thinks?

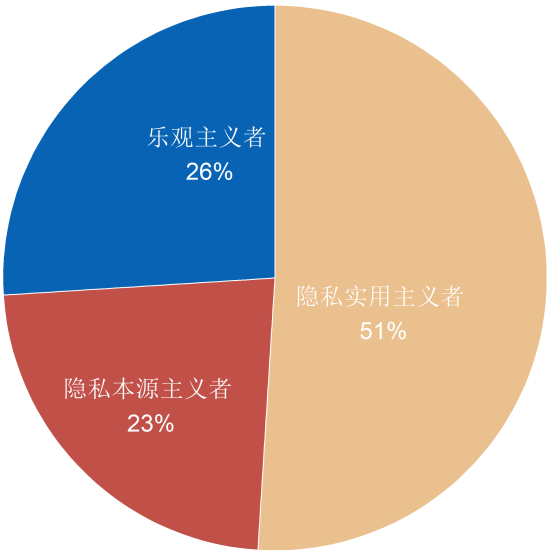


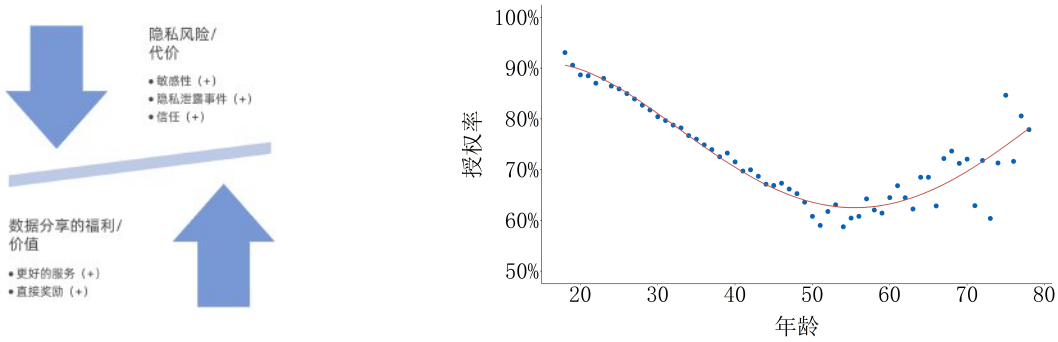
图 9: 消费者隐私态度分类

资料来源：全球数据驱动营销联盟 GDMA（2018）。

从消费者权益的角度出发，我们能更全面地理解消费者对待隐私的真正态度，以及他们在分享个人信息时的偏好。受到行为经济学家启发 (Kahneman and Tversky, 1984)，我们可以得到一个结论，当消费者分享个人信息时，他们既有可能获得福利，也面临风险 (Culnan and Bies, 2003)。这种共识后来发展为所谓的 “隐私权衡理论” (Culnan and Bies, 2003; Dinev and Hart, 2006; Laufer and Wolfe, 1977; H. J. Smith et al., 2011; Xu et al., 2009)。隐私权衡理论认为，在分享个人数据时，消费者通常会进行一个成本收益分析，将福利与潜在的风险进行权衡，如图10a所示。当预计的收益大于已知风险时，消费者倾向于披露他们的个人信息，反之则会拒绝。

研究支付宝小程序的发现与隐私权衡理论的观点相符。当消费者进行分享信息的决策时，信息的敏感度、隐私事件以及服务的质量都是他们计算的相关因素。更为重要的是，不同人群，或相同人群数字服务的经验的多寡，都会产生不同的权衡结果。整体而言，年轻人对数字服务最热情，一方面他们隐私信息相对较少，态度更加乐观，另一方面，与其他年龄组相比，他们对数字技术更加熟悉且从中享受到更多福利（图10b）。有趣的是，从 20 岁年龄组到 50 岁年龄组，随着年龄增长，授权率呈下降的趋势，50 岁以上又开始提升，似乎分享个人信息的风险在权衡比重中下降了。

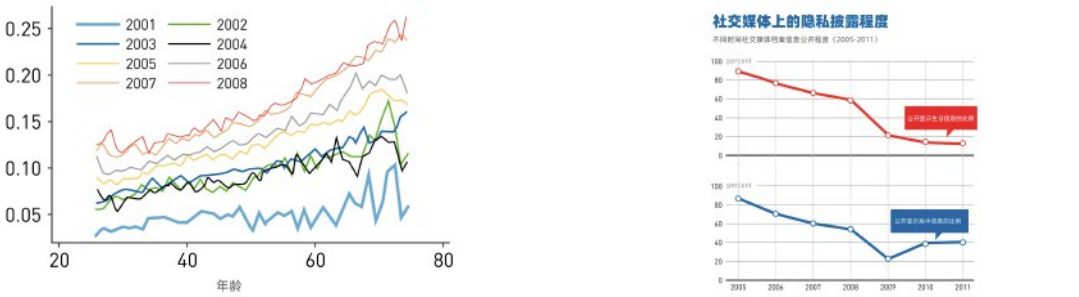
此前的研究也发现，隐私问题正变得越来越重要。Goldfarb 和 Tucker 指出，在美国，人们对隐私的担忧正在增加（图11a）(2012)。他们在 2001 至 2008 年投放了在线问卷，调查人们是否愿意披露自己的收入信息。结果显示，拒绝率每年上升 1.3%。所有年龄层都出现了相同的趋势。类似的是，2015 年 Acquisti 等学者的研究发现，从 2005



(a) 隐私权衡　　(b) 授权率与用户年龄

图 10: 隐私计算和消费者授权率

资料来源：罗汉堂。



(a) 拒绝分享收入信息的比例　　(b) 社交媒体上披露个人信息的比例

图 11: 不同时期个人信息披露习惯的变化

资料来源：(a) Goldfarb and Tucker (2012); (b) Acquisti (2015)。

到 2011 年，在卡耐基梅隆大学的 Facebook 网络里，公开分享自己生日和高中信息的比例在逐年下降（图11b）。

第 3 章 数据的价值

上一章的大数据分析表明，大多数用户愿意分享个人信息，从而享受数字服务。这自然引出了新的问题：数据到底给用户带来什么价值？为什么用户愿意分享数据？

在本章中，我们将探寻以上问题的答案，并从**连接**、**信任**、**决策**这三个角度探讨数据分享的价值。

3.1. 信息在数字时代的变革性意义

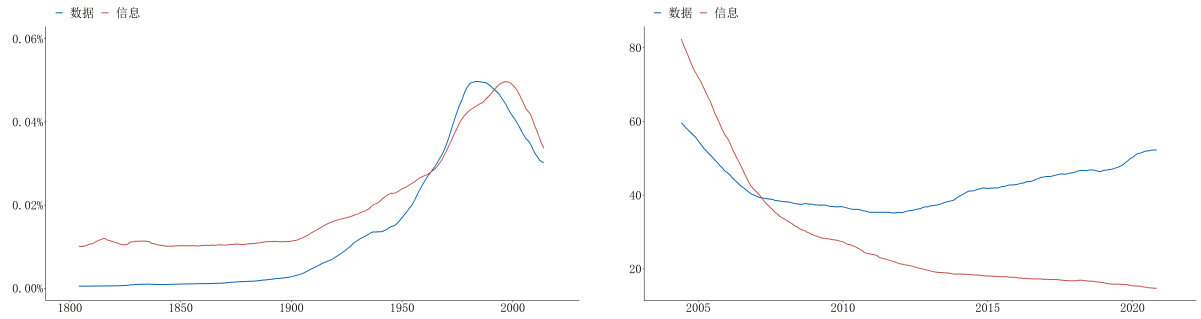
经济学家早就意识到，信息交流在经济活动中的地位不可代替。**哈耶克在这方面提出了两个关键论点：首先，不像普通的有形商品，做决策所需要的信息往往没有具体的存在形式，而且分散在各处，需要进一步生产和提炼；第二，为利用好分散的信息，社会面临最重要的经济问题，实质上是如何促进信息的收集和交流**（哈耶克，1945）。

这听起来像是简单的常识。但由于近些年隐私泄露、身份盗窃和网络犯罪屡见不鲜，造成的危害也显而易见，随着隐私问题的辩论日益激烈，人们关注的焦点往往集中在个人信息交换的负面影响上，而忽略了常识。我们不应忘记，信息分享不仅给每个人带来了福利，对整个社会的进步更是至关重要。

一些开创性的经济研究出现于 20 世纪 70 年代和 80 年代，其结论表明，有限信息和不对称信息将阻碍市场有效率地配置商品和服务，阻碍自愿和互利的贸易，并导致市场失灵，影响宏观经济政策的效力，扭曲投资和消费决策，造成失业（Akerlof，1970；Milgrom and Stokey，1982；Myerson and Satterthwaite，1983；Phelps et al.，1970；Pissarides，2000）。这说明，信息分享和扩散会影响到人类协同合作的水平。通过巧妙的市场设计和工具，能够在一定程度上缓解信息不对称所带来的效率扭曲。这些工具包括信号理论（Spence，1973）、筛选理论（Vickrey，1961；Mirrlees，1971），以及更通用的机制设计理论（Dasgupta，Hammond and Maskin，1979；Green and Laffont，1981；Myerson，1981；Maskin，1983，1999，2008）。

始于 20 世纪 40 年代的信息革命，从根本上改变了人们生产和使用信息的方式。1946 年，“数据”一词首次被用来表示“可传输和可存储的计算机信息”。20 世纪后半叶以来，“数据”和“信息”两个词在全球的使用大幅增长（图12a）。进入 21 世纪，新一轮数字化的革命的到来使得“数据”一词的使用进入一个新的阶段。以 Google Trends 为例，自 2007 年以来，“数据”的全球搜索频率已经持续高于“信息”（图12b）。

经过信息革命，数据的形成、生产、存储和通信成本空前下降，这引爆了数据使用的热潮。摩尔定律揭示了一个趋势，就是处理器的性能每隔两年翻一倍，计算成本也相应下降。其他计算领域的进步等也加速了数据增长。例如，云计算提供了处理共享数据的能力，从而实现更高效的业务计算。人工智能和机器学习技术的进步，利用光传输数据的硅光子学等新技术彰显了数据处理速度提升的巨大前景。今天存储和传输数据的边



(a) 1800 年以来谷歌图书收录的所有在美出版英文书籍中出现“数据”和“信息”的比例。

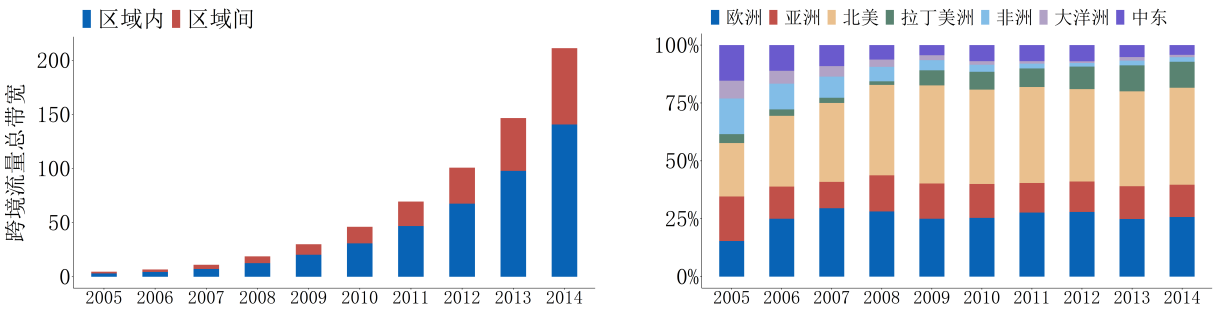
(b) 2004 年以来全球在谷歌搜索上“数据”和“信息”搜索率的相对变化趋势。

图 12: 20 世纪后半叶以来，全球进入“数据”和“信息”的时代

资料来源：谷歌图书 N 元语法查看器（Google NGram）和谷歌趋势（Google Trends）。

际成本已经降到了几乎为零的程度。此外，“数据”这一术语已演变为“大数据”，体现了数据处理的空前规模、维度和速度。

事实上，全球产生和存储的数据总量从 2009 年的 0.8 ZB（万亿 GB）增加到 2018 年的 33 ZB，并预计在 2025 年达到 175 ZB（Reinsel et al.，2018）。就种类而言，由于数字化成本收益比极高，基本上所有种类的信息都可以数字化。今天数据生成和流动的速度之快在过去很难想象。一条消息在推特上发布后，立即会被记录在推特的数据架构中，并在几微秒后发布到用户的时间轴上。全世界成千上万的用户在推特上发布信息，从而产生了不间断的海量实时数据流。大数据中的“速度”也意味着信息实时生产和分享的数量之多和维度之广。数据流动的速度持续加快，全球互联网的数据流动速度在 10



(a) 区域间与区域内信息流动增加

(b) 各大洲信息流动情况

图 13: 全球数据流动呈指数型增长

资料来源：联合国贸易和发展会议（2019）。

年内增长了 20 多倍，从 2007 年的每秒 2000GB 增加到 2017 年的每秒 4.6 万 GB（联合国贸易和发展会议，2019 年），相当于每秒传输美国国会图书馆全部藏书数据量的 4 倍。这已成为一种全球现象：数据在区域内和区域间持续流动，流量在世界各地都呈指

数级增长（图13）。

如前文所述，“数据”并不等同于“信息”。“数据”作为数字化的记录，可以看作是信息的载体或媒介，但数据不一定包含信息。例如，添加一组随机算法生成的数据不能传达出任何信息。除此之外，生产（或观察）实体或虚拟世界的原始数据并不简单，还要对数据进行处理，才能获得有价值的信息（图14）。经济学家 Singh （1999）指出，密码学通过加密算法将信息深埋在数据中，这正是利用了数据和信息之间的巨大鸿沟。虽然数据本身可以被“公开”并因此可免费获得，但只有拥有加密“密钥”的人才能够从中提取信息。读者应了解这种差异，我们在报告中将“信息”和“数据”互换使用，仅在必要时指出两者之间的区别。

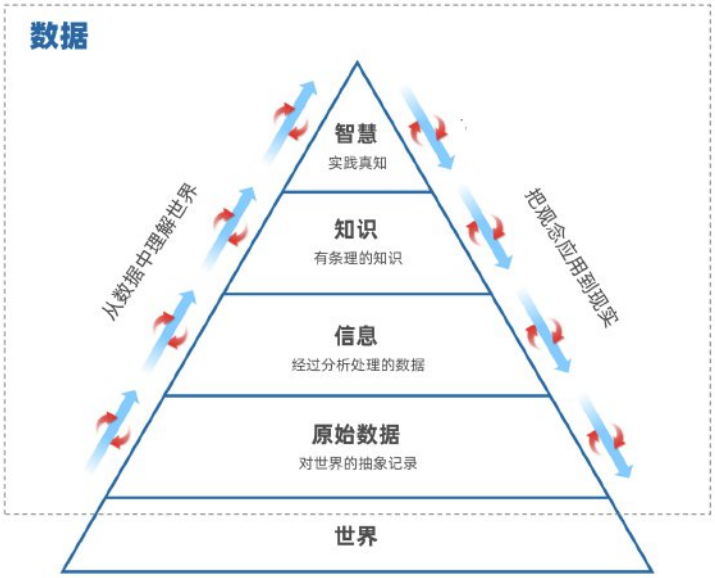


图 14: 知识金字塔

资料来源：罗汉堂【参考 Kitchin（2014）和 Boisot and Canals（2004）】。

3.2. 数据在数字时代的价值

数据只有在使用时，在经济生产和社会活动中流动时产生价值。随着人们对数据的处理、传输速度大大提高，人类协作的三个基石——**连接、决策和信任**正在信息革命中发生着根本性转变。以下讨论将对这三大要素进行详细分析。

3.2.1 数字化连接: 普惠性参与和协作达到前所未有的水平

正如我们在《新普惠经济：数字技术如何推动普惠性增长》（罗汉堂，2019）指出，数据分享加强了连接。由于数据非常容易生成和分享，普惠性连接达到了前所未有的高

度，这重塑了市场以及人们协作生产和消费活动的方式。

亚当 · 斯密在《国富论》中指出，“劳动分工受市场范围的限制”。贸易一直被广泛定义为“重力模型”（gravity model）：贸易的数量和频率与距离呈负相关，与市场规模呈正相关，距离越远，交易越少。而在今天，在互联网技术和全球通信、物流体系下，距离已经不是制约交易的重要因素。

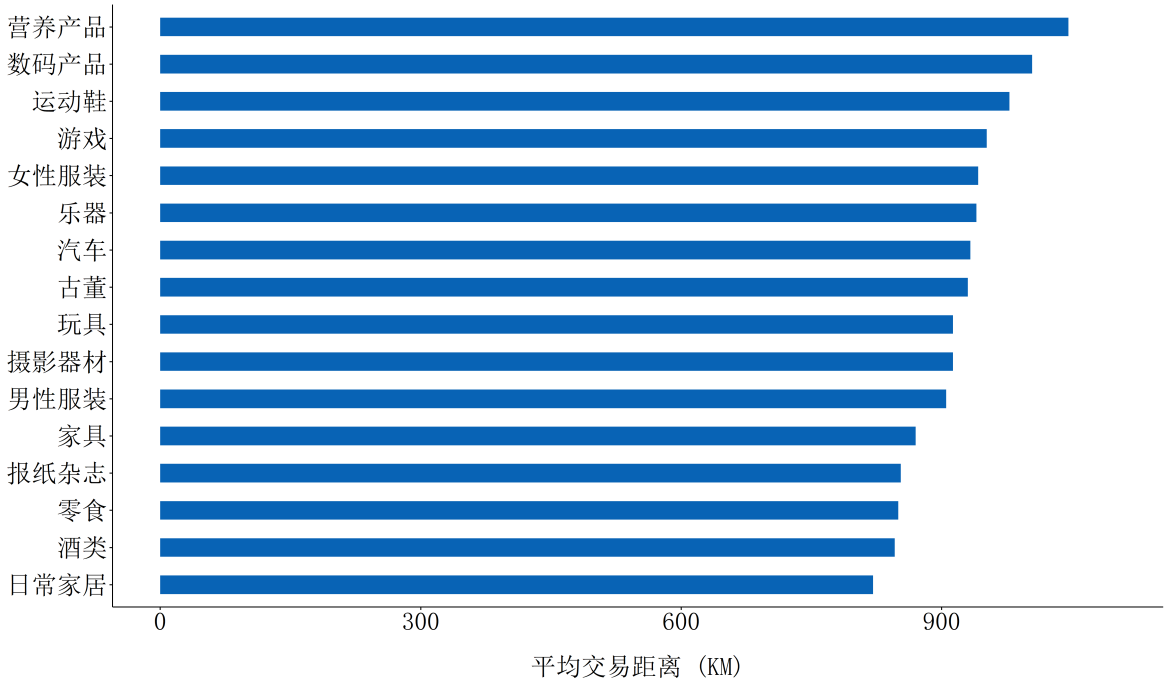


图 15: 淘宝与天猫不同品类物品平均交易距离，2018

资料来源：罗汉堂（2019）。

注：该样本包含 2018 年随机选择的一天中进行的所有交易。交易距离定义为发件人和收件人所处省份省会城市之间的距离。

以中国最国际化的城市上海为例。即使在最热门的商业区，超过 80% 的线下客户仍处于商圈中心 10 公里的区域内[1] 。如果把距离拉得更远，买家和卖家根本感知不到对方。他们对商品和服务的品种、质量、价格，以及客户需求、卖家信誉等细节缺乏准确的信息。

在全球范围内，阿里巴巴和 eBay 等双边电子商务平台描绘了一幅让亚当 · 斯密无法想象的画面。在线交易极大拓展了贸易的范围、深度和广度。在电商平台上，除了生鲜食品，买家和卖家之间成交的平均距离接近 1000 公里，超越传统线下市场服务范围两个数量级，“重力模型”可能已被打破（见图15）。从连接买卖双方的情况看，淘宝月度活跃买家超过 7.2 亿，卖家有超过 1000 万家小微和初创企业，其中约一半的创业者是女性。产品变得极大丰富，消费者在线上可购买 10 亿种以上的商品和服务。

这种远距离的交易之所以能发生，原因在于信息流动大幅提速，消费者与生产者匹

1 参见文章，大数据告诉你，上海市高峰人口有 3000 万。

配效率明显提高，物流体系也因为信息流动、交通改善等因素变得更为快捷。由于客户有数十亿种商品和服务可供选择，所以根本不可能搜索到所有感兴趣的产品或服务，生产商也无法接触到所有潜在客户。如果说传统市场的主要障碍是缺乏信息，那么数字时代的新障碍就是信息太多——信息超载。在这种情况下，低效信息俯拾皆是，人们更需要有价值的信息，因此为买卖双方牵线搭桥的有效机制至关重要。这就是“大数据”的意义所在。

3.2.2 数据分享优化决策

海量、多种类的数据，再加上快速连接，让无数客户和生产商做出更明智的决策，从而促成更快速、有益的产品创新，更具创新性的销售和服务，以及新商业模式——或者说产业组织的新形式，而这些在以前根本不可能实现（罗汉堂，2019）（见第六章对熊彼特式竞争的讨论）。

除了使用搜索和店铺列表等传统工具外，电商平台利用越来越多的推荐系统，能更有效率地帮助消费者找到自己想要的产品。而这个推荐系统则依赖一些大数据信息，电商推荐系统依据的消费者数据包括购买历史、搜索活动和个人特征（但不是个人的具体身份信息）等，匹配的推荐通过算法完成，因此供应商可以“感知到自己的客户群，但不知道他们具体是谁”。尽管这些匹配算法非常有效，但买卖双方都是最近几年才开始探索其潜力。到目前为止，只有少量相关数据被用于帮助匹配买家和卖家、用户和供应商。

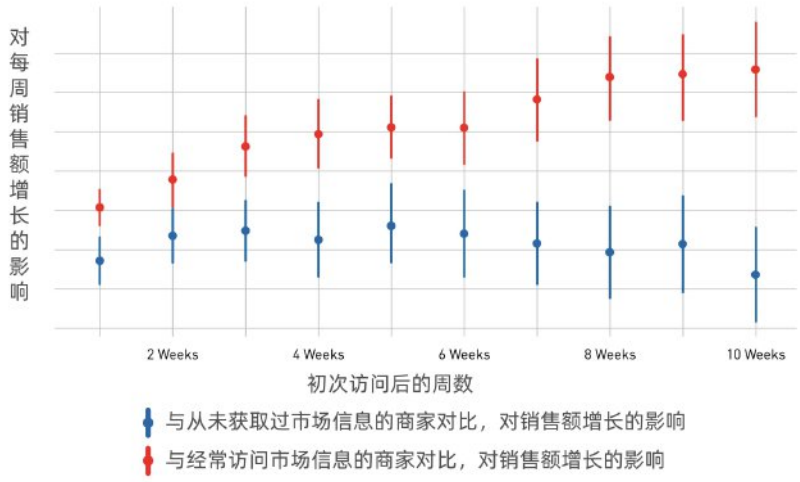


图 16: 是否使用数据服务对销售的影响

资料来源：罗汉堂。

注：与从未访问过市场信息的商家相比，最近开始获得此类信息的商家经历了后续可观的销售增长；与经常访问市场信息的商家相比，新近使用信息服务的商家的销售增长仍然是积极的，但增长不那么强。

中小微企业和个人，原来很难有效寻找到最合适的客户或者产品信息。数据流动增强后，消费者和生产商都能做出更明智的决策[2]。例如，阿里巴巴的生意参谋为所有网店店主服务，提供各种信息分析工具，例如他们自身的历史业绩分析、市场趋势以及潜在消费者需求等。新订阅用户（其中大多数是中小微企业）通常会在使用的第一周内经历销售增长的大幅跃升。在接下来的十周内，已订阅和未订阅组之间的业绩差距愈发明显（图16）。淘宝、京东、亚马逊和 eBay 等数字平台都提供类似的信息服务，帮助卖家（无论规模大小）做出商业决策。“大数据”使数据成为一种服务，为中小微企业配备了此前大公司才能获取的多种复杂分析工具。我们也可以称之为智能普惠化。

在金融领域，大数据让小额贷款的规模化成为现实，而这在以前根本不可能实现。

在金融领域中，“了解你的客户”（Know-Your-Customer ，简称 KYC）从来都扮演了重要角色。从历史上看，正是因为缺乏信息，大多数企业贷款都是以抵押为基础，而大多数中小微企业由于缺乏抵押品而无法获得贷款。这造成了中小微企业的巨大资金缺口，如何为这些企业提供有效的金融服务一直是世界性难题。

金融科技的出现扭转了困局。金融科技贷款利用大数据，服务于担保额度低、但具有高增长潜力的中小企业，而传统金融中介更多依靠信息不敏感的抵押品提供贷款，这样的模式主要为有抵押品的大公司提供贷款。自 2011 年以来，以网商银行、微众银行、京东金融等为代表，中国开始利用大数据向中小企业和初创企业提供无抵押贷款，可以做到“310“模式：申请不到 3 分钟，1 秒即可获得贷款，0 人工干预，而且整体坏账率可控。这个模式也越来越成为有大数据能力的银行的标准。

由于有企业经营的数据信息，小企业无须提供实物抵押即可获得融资的服务，这克服了普惠金融迄今难以逾越的障碍。得益于大数据和金融数字化，用 Holmström 的话来说，信息已成为新的“抵押品”，帮助许多初出茅庐的企业家取得了成功（Holmström，2018）。如图17所示，Hau 等（2018 年）经济学家研究了基于信用评分的贷款效果，发现获得贷款的中小微企业销售增长明显高于没有获得贷款的中小微企业。

Berg 等（2020）学者同样研究了数字足迹对信用评估的影响。“数字足迹”是消费者在注册或浏览网站时在网上留下的信息。他们发现即使是简单的数字足迹信息也可以成为传统征信机构评价信息的有益补充。银行结合使用征信机构和数字足迹信息，已经能够将违约率降低大约三分之一。因此数字足迹可以增加没有银行账户的人口获得信贷的机会。

全球有多达数十亿的贫困人口，不仅长期缺乏获得贷款的机会，甚至缺乏金融账户。相比之下，中国在十年的时间里转变为一个拥有超过 10 亿用户的移动支付国家，值得注意的是，中国基本消除了贫困人口，很多人在这十年里也获得了借贷服务。

在中国，在大数据三个 V 特质的推动下，数字支付系统变得可靠和可持续，金融机构得以快速地为不同需求的大量借款人提供贷款。当用户用传统的实物信用卡或借记卡进行支付时，我们很难知道刷卡人的身份和信用信息。相比之下，用户使用移动支付

[2] 参见 Veldkamp 和 Chung （2019），Carriere-Swallow 和 Haksar （2019），以及 Jones 和 Tonetti（2020）总结了对数据在经济总量中所起作用的最新研究。

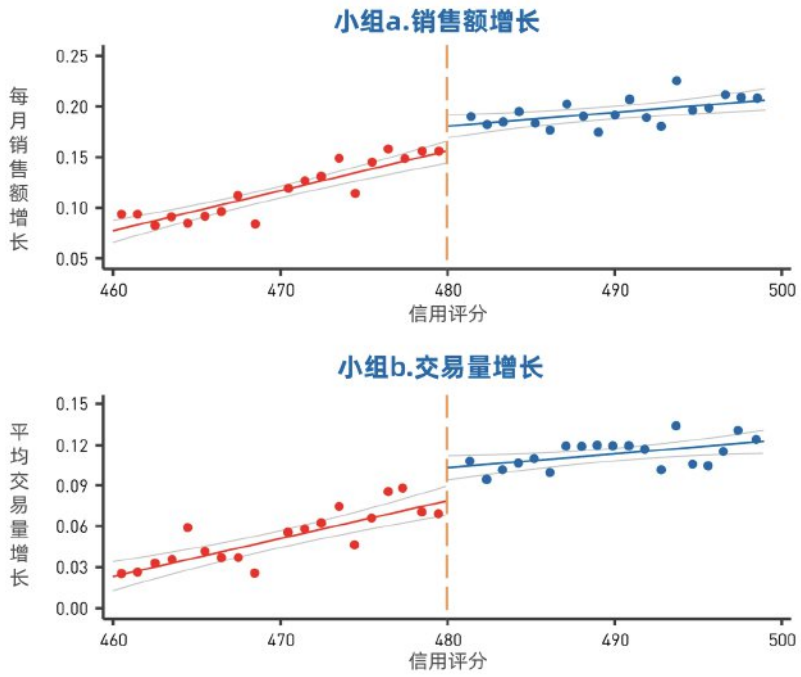


图 17: 基于信用评分的分化贷款政策带来了企业营收的分化

资料来源：Hau 等（2018）。

注：数据采样期为 2014 年 9 月至 2016 年 7 月。

时，相关的位置、时间、消费习惯以及买卖双方身份等额外信息都会随之产生。移动支付算法会基于这些附加信息，来评估支付人是否拥有账户，账户中是否有足够的资金，以及此人是否做出了不同寻常的支付行为（风险信号）。根据中国人民银行的数据，信用卡和借记卡的平均欺诈损失率为 0.02%。但以支付宝为例，移动支付的欺诈损失率不到 0.0005%。有了实时的风险评估，用户（包括没有银行账户的人群）现在享受到安全便捷的移动支付服务，而这反过来又增强了用户数据的实时交互。

在数字时代，个人信息的分享已经成为有效匹配供需的一个关键环节。比如用户访问淘宝时，他们通常会看到一系列推荐产品。自动匹配算法会基于相关个人数据，来进行个性化推荐。**那么在这一过程中，个人数据的使用有多重要？如果在商品推荐时不使用个人数据，会发生什么情况？**为回答这些问题，Sun 等学者（2020）对超过 62 万名用户进行了大规模随机实验。在实验中所有参与者被随机分配到实验组和对照组。在对照组中，产品推荐的匹配算法保持不变，但实验组不能使用个人数据进行算法推荐。这样，实验组和对照组之间的比对有助于量化个人数据的价值。

实验结果是惊人的。使用了个人数据的一组，客户页面浏览量大致均匀地分布在头部商品和长尾商品之间（图18）。没有使用个人数据的一组，推荐明显集中在少数几种商品上，排名前 1000 位的商品几乎占到了所有曝光量的 90%。

实验还显示，随着可供选择的产品种类的减少，消费者的参与度会大大降低，导致点击率下降 77%，商品浏览量下降 33%（图19）。由于展示给消费者的产品吸引力并不强，搜索量会大幅增加。

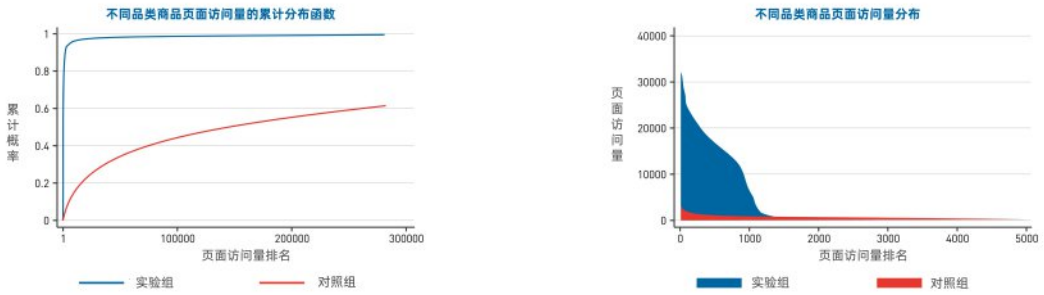
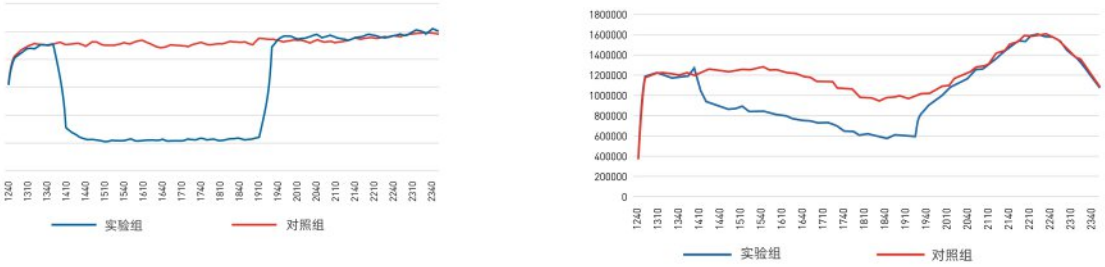


图 18: 是否使用个人数据推荐商品会带来很大差异

资料来源：Sun（2020）。

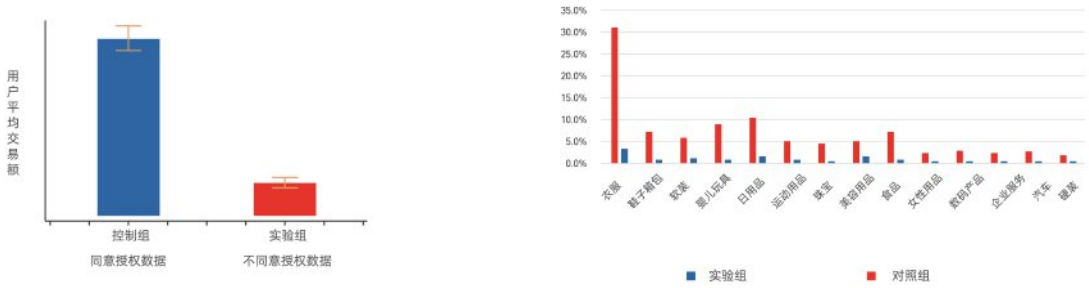


(a) 用户对推荐商品的点击率下降 77%　　(b) 用户对首页推荐产品查看次数下降 33%

图 19: 用户点击率和产品访问率

资料来源：Sun（2020）。

更严重的是，由于匹配效率低下，客户需求无法被很好满足，最后交易频次下降高达 81%。推荐产品数量减少以及客户参与度降低，还会进一步导致市场成交量大幅下降（图20）。今天，越来越多的在线消费很大程度上依赖个性化推荐。如果不依据个人数据，自动推荐功能会被大大削弱。供应商只有“了解”客户，才能满足客户的需求。如果对个人数据分享采取过于保守的态度，虽然降低了隐私风险，但消费者却丧失了推荐服务带来的价值。



(a) 总商品交易量（GMV）下降　　(b) 剔除个人数据对电子商务中所有行业

图 20: 是否有个性化推荐会带来交易效率的很大差别

资料来源：Sun（2020）。

政策机构曾经担心，“公司可以利用大数据将缺少服务的低收入人群排除在信贷和就业机会之外”（联邦贸易委员会，2016）。各种实证证据和实验结果表明，阻碍个人数据的分享和流动，可能会更多地让更需要支持的群体受到伤害，其中包括平台的新用户、低收入地区的居民、女性，以及身体或其他方面有缺陷而需要在线医疗、金融和其他服务的人群。为高效地服务客户，生产商和供应商需要“了解你的客户”（KYC）。虽然这是常识，大规模随机实验也清楚地表明，当个性化推荐系统中的个人数据流被切断时，所有参与者，尤其是更需要支持的群体，都会受损，从而带来巨大的社会福利损失。所以，在保护适当的前提下，让数据流动畅通，所有的参与者都是数据交互的受益者。

3.2.3 数字化建立信任

Akerlof 在他 1970 年的经典文章《质量不确定性与市场机制设计》中，曾经用“柠檬市场”生动形象地说明，服务经济中有很大一部分经济活动会因信息不对称而消失。这是因为消费者和生产者信息不对称，消费者缺乏对产品的信息和信任，只愿意选择低价产品，从而劣币驱逐良币，赶走了好的服务商，只剩下质量不好的“柠檬”，随之恶性循环，直到整个市场消失。用经济学的话来说，柠檬市场现象说明，可信信息的缺乏可能导致“逆向选择和道德风险的代理问题”。在数字时代，实时数据的流动和使用是经济活动中的关键一环，能够遏制多种机会主义行为，让可信的参与者受益。比如实时使用的数据能够提高普惠金融服务的可及性和覆盖面，实时数据结合机器学习和人工智能

算法，可以提供越来越准确和及时的评估、建议，从而让买卖双方都能受益。

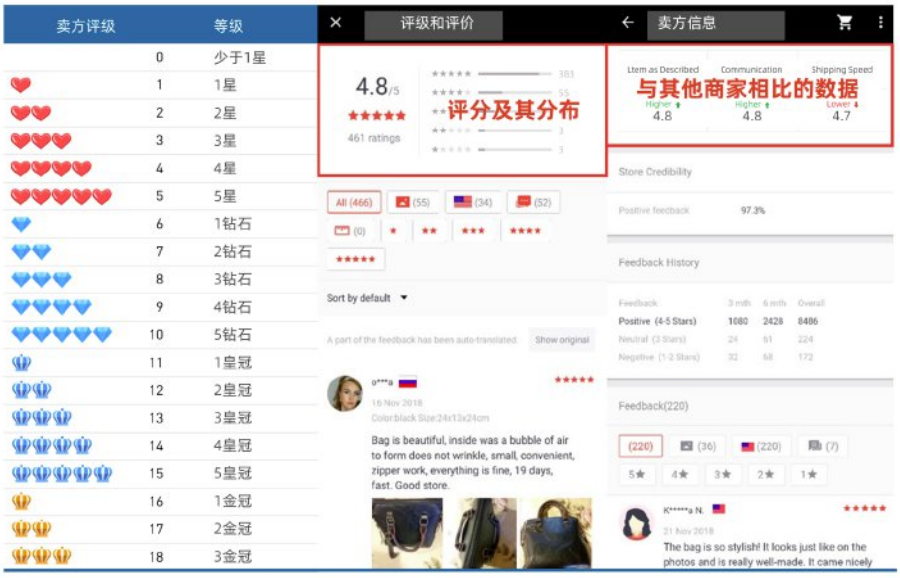


图 21: 淘宝通过消费者评分建立商家声誉

资料来源：罗汉堂。

所以数据的分享可以帮助建立信任。在线市场中，全球数以十亿计的人们彼此间达成交易，如何能让买卖双方像在本地市场面对面交易一样互相信任？这是运营在线市场的关键。解决办法就是数据分享。通过在线数据分享，客户可以对商品和生产商评级。因为所有参与者都能看到这类评级和评价，生产商会格外注重建立声誉。通过这个数据分享机制，所有善意的参与者都能从中获益，这与传统的“柠檬市场”形成了鲜明对比。评级系统给予买家和卖家通过信息分享构建信任的权利。围绕建立卖家的长期声誉，为平台产生高质量、可持续的卖家创建了一种激励机制（Tadelis，2002）。数据不仅让买家受益，也让高质量、有回头客的卖家更好地将自己与低质量、无信誉的卖家区分开来，建立品牌意识，为长期的销售表现带来了动能。

淘宝十多年的评级机制清楚地表明，通过数字技术的应用（图21），信用是如何有效并准确地建立和升级的。淘宝对卖家采用的是“红心-钻石-皇冠”评级系统。卖家积累消费者的好评，获得了红心。五心卖家升级到钻石，五钻卖家则升级到皇冠。评级系统使用的数据来自消费者，他们愿意分享购物体验和售后产品使用体验。这样，高质量的卖家可以通过信息分享脱颖而出。此时，即便不能在每一笔交易中完全消除“柠檬市场”效应，信息不对称带来的影响也会小得多。

通过数据分享建立商家声誉有多重要？一个衡量方法是观察卖家评级提升时会出现什么情况。由于商家的业务基本面是长期不变的，而买家评级是分散自发的，因此可以通过数据统计来评测声誉变化的价值。如图22所示，我们发现在评级提升后的一个月里，

卖家的销售额通常会有显著增长。这既说明声誉很重要，也表明评级提高有较大的价值。尤其当评级从零到一颗红心，从五心到一钻，从五钻到一冠的时候，销量增幅最大。因为信用升级，销量提升了。（罗汉堂，2019）

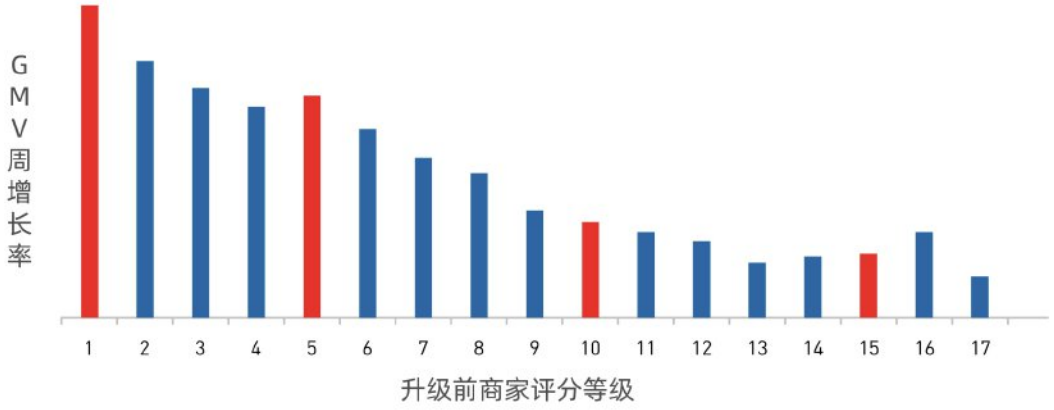


图 22: 评分上升后 GMV 的周增长率，2017

资料来源：罗汉堂。

注：评级分布选取随机样本中 2017 年销售额为正的卖方

我们常常用“信息流”“资金流”和“商品流”来衡量经济活动。信息流动是所有基于协同的经济活动中不可或缺的部分；没有信息流动，资本和消费品就不能从一方流向另外一方。用哈耶克的话来说，经济问题就是信息问题。信息将人们连接在一起，使生产商知道如何为客户服务，建立信任，并做出更明智的决定。数字革命将数据总量、种类和速度提升到了前所未有的水平，大数据成功改变了消费者和生产者之间的联系，进一步增强了买方和卖方之间的信任，并促进了更好和更快的决策。

当然，随着数字技术的不断进步，我们也需要解决随之而来的隐私和安全风险。我们将在下一章阐释这个话题。

第 4 章 隐私风险、隐私保护和数据安全技术

现代隐私观念起源于 19 世纪 80 年代美国的法律实践。著名的美国最高法院法官 Louis Brandeis 将隐私称为“独处权”（Warren and Brandeis，1890），认为它是人类尊严、自由、能动性和尊重的基础。如今，隐私权在许多宪法和国际条约中已经被视为一项基本人权，如《世界人权宣言》、《公民权利和政治权利国际公约》和《美洲人权公约》，中国的《民法典》也明确规定自然人享有隐私权，是其人格权的一部分。对隐私的重视和保护已成为全球各界的广泛共识。

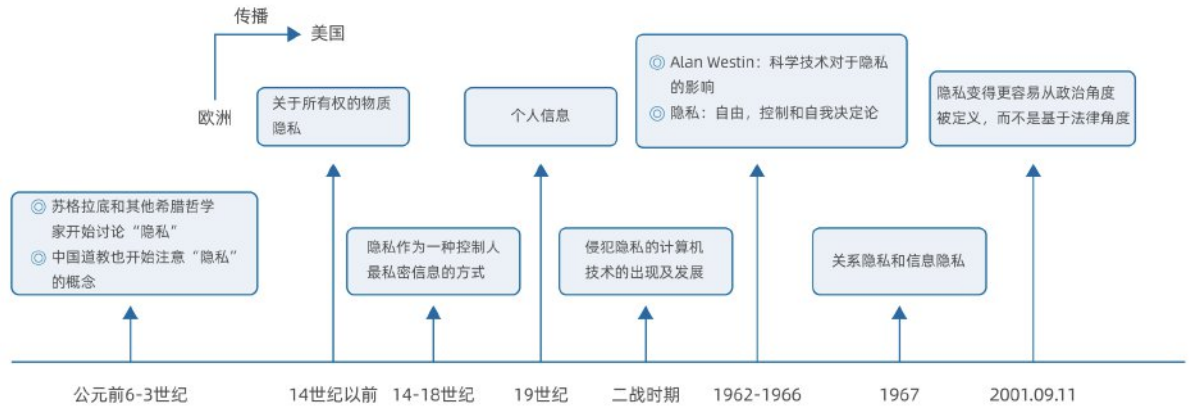


图 23: “隐私”概念的演变

资料来源：Ballard （2013）；Holvast （2007）；Dempsey （2019），罗汉堂概括整理。

对隐私的在意不是一个新现象。(图23) 早在 Brandeis 之前，几乎所有古代文明以及宗教著作，都提到了个人和群体隐私的需求 (Banisar and Davies，1999)。亚里士多德将人的生活首先区分为公共空间和私人空间，个人对私人空间应当享有更强的控制。《礼记》中也有“将上堂，声必扬”的论述，提醒不要悄悄进入别人的隐私空间，教育人们要考虑到他人的隐私。虽然隐私的涵义在不同文化、背景和环境中有所不同，包括“控制”“保密”“亲密”“尊严”“自主”“信任”和 Brandeis 的“独处权”等，但这些正说明了隐私是人类的基本和普遍需求之一。甚至如经济学家 Volio 在其 1981 年的研究中所强调的，“从某种意义上说，所有人权都是隐私权的一个方面”。

不同时代隐私保护的侧重和迫切程度有所不同。19 世纪西方开始重视隐私权的保护，背景正是第二次工业革命带来的电话、电报的广泛应用，产生了信息的远距离快速

传输，也带来了私人空间中的隐私信息泄露的风险。随着信息技术的发展，越来越多的信息可以通过声音、影像的方式被记录、复制、传播，这推动了 20 世纪 80 年代以来的信息保护实践。伴随着互联网，尤其是移动互联网以前所未有的速度进入人们的生活，信息交换和分享的维度、速度和量级都前所未见，相应也大大增加了隐私被侵犯、信息被泄露的风险，隐私保护成为一个全球性挑战。

回顾历史，保护隐私的制度安排也有共性，即从来都不是把隐私简单界定为一项不可剥夺的权利，而是将“隐私”视为控制信息和从自有信息中获得福利的权利（Schwartz，2004）。这种思路的背后是认识到信息分享的价值，认可消费者对涉及隐私的信息的控制权，因而允许消费者放弃部分隐私，以便享受信息分享带来的好处。在数字时代，这意味着个性化营销体验、定制化的金融服务、医疗保健、教育，以及便捷的社交网络。换句话说，为了保护好隐私，而不是流于形式，最有效的做法是将隐私视为一种可交换的商品，使参与者有权选择通过让渡部分权益得到好处。正如著名美国法学家 Richard Posner 指出，太多隐私倡导者将“避世”——即大法官 Brandeis 所说的“独处权”——与“保密”，即控制信息的权利混为一谈（Posner，1979）。

4.1. 数字时代隐私风险源于何处？

数字时代在放大了信息分享带来的好处的同时，也增加了隐私风险。数字经济的特征是把越来越多多维度的、碎片化的、实时的小数据转化为“大数据”，在此基础上提供各类线上服务，让消费者和商家都得到好处。但因为数据的广泛使用，在数据周期的每个阶段，从数据收集到存储、分析、使用，到数据清除阶段，都存在隐私泄露和数据安全的风险。

以**数据收集**过程为例。看起来，数据收集只要在个人知情和同意的前提下，就没有问题。但在实践中，保护个人免受过度或未经授权的数据收集是一项艰巨的挑战。道高一尺魔高一丈，黑客和网络钓鱼者会用尽浑身解数开发出新的手段和技术来不当获取数据。

网络钓鱼是对一类骗取个人数据行为的统称。通过模仿某个值得信任的实体，与消费者的个人设备或日常服务取得联系，为的是骗取个人数据。常见的网络钓鱼方法包括带有链接的垃圾邮件、浏览器中的弹窗，或精心编制过话术的电话钓鱼，后续常常伴随着金融盗窃或诈骗。根据卡巴斯基实验室的数据，仅在 2019 年第一季度，其 12.1% 的用户遭到攻击，而其反网络钓鱼系统阻止了超过 1 亿次将用户跳转至诈骗网站的尝试[1]。

2018 年，一家上市公司北京瑞智华盛被曝非法收集了 30 亿条个人数据记录。该公司与各地网络运营商合作，以精准广告营销为名，获取了远程登录其操作系统的权限。然后，公司将数据收集程序嵌入运营商的系统中，自动收集客户数据，例如包含账户和密码的 cookies，以及存储在本地服务器中的数据。有了这些数据，瑞智华盛就能够在多个平台上登录众多客户的账户。该公司以每个用户 0.5 元人民币左右的价格为许多社交

[1] 参见卡巴斯基实验室数据，Spam and phishing in Q1 2019。

网络平台提供违规的营销服务，获得巨额非法利润。部分泄露的个人数据还被用于金融诈骗，带来了更大的损失。这个案例还包括不法分子对网络基础设施的肆意攻击，也警示我们要对数据的全流程风险更加关注。

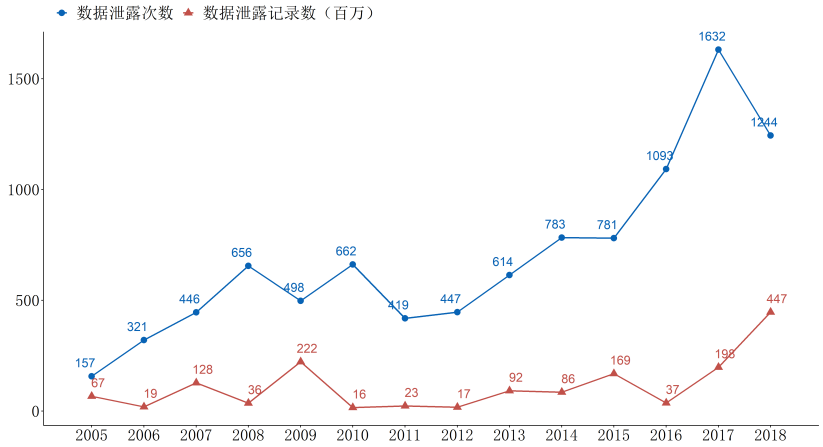


图 24: 2005-2018 美国年度数据泄露的次数和记录量（百万）

资料来源：2020 Annual Data Breach Report by Identity Theft Resource Center。

网络追踪技术本意是通过追踪消费者的浏览内容，帮助网站背后的服务商发现不同客户的真正兴趣，从而有针对性地提供个性化的服务（Hoofnagle et al.，2012），但是用户通常对在线广告背后的技术一无所知（Smith et al.，1999）。网络钓鱼者和黑客可以利用这种认知上的不对称，从受害者那里窃取信息或以其他方式欺骗受害者。由于这些都发生在浏览网页的后台，受害者对他们的数据被收集的程度、由谁收集以及以何种方式收集一无所知（McDonald，2010）。此外，对许多人来说，了解这些技术的门槛太高，实际上没有办法选择摆脱追踪。

接下来分析**数据存储**阶段。个人数据通常存储和汇集在本地服务器或云端，都面临着被恶意攻击的潜在风险。多份报告显示，在 2016 到 2018 年期间，发生了涉及 11 TB 数据记录的泄露事件[2]，涉及的人数超过十亿。以 Facebook 为例，2018 年 9 月的一次攻击，使得 5000 万用户的账户面临威胁，这是该公司历史上最大的一次泄露事件。随后，在 2019 年 12 月又发生了另一次数据泄露，超过 2.67 亿用户的信息被黑客论坛的在线数据库获取。其中包括用户的 ID、全名和电话号码。

Facebook 及其用户并不是唯一的受害者。其他公司及其用户也经历过数据泄露，而且形势越来越严峻（图24）。在美国，数据泄露事件的数量从 2014 年到 2017 年翻了一番。例如，威瑞森数据泄露调查报告（2015）统计了 2100 多起案件，2014 年泄露记录超过 7 亿条。截至 2021 年初，全球范围内泄露信息超过一亿人的恶性事件已经达到 29 起。

在**数据使用**阶段，即使在合法收集之后，也可能出现将数据挪作他用、甚至转卖数

2 参见纽约时报报道，Facebook Security Breach Exposes Accounts of 50 Million Users。

据的行为。最有名的案例之一当属 2018 年“Facebook-剑桥分析公司”数据丑闻，损害了数百万 Facebook 用户的利益。剑桥分析公司开发的应用程序“这是您的数字生活”（This is your digital life）要求 Facebook 用户同意完成一项学术调查。用户同意后，该程序收集了用户社交网络中所有信息，但是这些数据最终被用于政治目的。该丑闻导致美国联邦贸易委员会对该公司处以 50 亿美元的罚款，这是迄今为止全世界最大一笔罚款，同时对该公司实施了严格的新隐私法规[3]。这一事件严重损害了 Facebook 的声誉。

最后，我们来看**数据清除**阶段。由于“被遗忘”的权利是隐私保护的一个重要方面，对历史数据的清除也是消费者的重要需求。一些搜索平台似乎在这方面取得了重大进展。比如谷歌、必应等浏览器可以要求追踪技术清除用户浏览历史。同时一旦某个设备中某条信息被删除，与其同步的所有设备也会清除该信息。另一个例子是各种浏览器开发的“隐私模式”浏览功能，在这一模式下消费者的浏览记录不会被追踪，尽管可能带来一些不便，例如自己也无法查看浏览历史，但是这类服务的推出给了不同消费者更多的选择，受到市场的好评。

4.2. 隐私工程化和隐私加强技术

我们每个人在使用数字服务的时候，都在有意无意地让渡部分个人隐私。通过选择点击“购买并保存”的按钮，把自己的购买记录保存下来，享受了记录信息的便利，虽然不一定非要这样。通过允许移动应用程序访问地址、浏览历史等个人信息，我们能够获得更精准的服务。另外，黑客和网络钓鱼者永远不会消失，他们将不停“开发”新的方式来骗取消费者的信息。

如何通过法规定义和保护好隐私权，已经越来越为社会所关注。因为数据交互是经济协同的基础，法规需要发展到哪个程度，才能既保护好隐私，又能支持数字经济的发展，最终造福整个社会，是一个很大的挑战。另外，无论法规如何健全，都需要落实到行业和企业行动中去。我们在下面专门讨论在企业层面如何基于法规和原则做好隐私保护和数据安全。

历史表明，新技术一方面会带来新的挑战，也会带来解决方案。与新药研发类似，新的技术能够通过限制数据分享中潜在的“副作用”来提高数据分享的安全性、透明性和可持续性。重要的是，在数字技术提供的各种福利和保护个人隐私之间实现适当平衡，尚无证据表明严格监管和巨额罚款是唯一或者最佳方式。另一方面，我们不可能回到避世的与外界隔绝的状态。只要有人类协同，个人隐私就永远不可能得到绝对的保障。而一味地依靠事后的惩罚来打击隐私侵害行为，会耗用太多本可用于预防其他类型犯罪的资源，损失了本可实现的更高层次的个人发展和社会进步水平。

隐私保护的关键在于用好数字技术，开发出更强的保护机制和更有效的保护技术。比如数字支付系统通过利用多维信息、实时风险甄别，以及人工智能算法，让支付中的

[3] 参见 FTC 媒体发布，FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook

欺诈行为大大降低。我们接下来主要介绍在隐私保护实践中近些年的两个有潜力且互补的核心方向：**隐私工程化（“隐私设计”）和隐私增强技术**。

隐私保护工程化正成为数字时代对企业的一项核心要求。许多科技公司都已经开始践行“隐私设计（Privacy-by-Design）”的方法。隐私工程化将隐私保护的法规和“用户导向”的原则引入到软件、服务设计和使用的各个环节中，将隐私保护前置，从产品和系统设计的初始阶段就考虑到如何解决隐私保护问题。隐私工程化包含两个部分，首先是软件的设计中加入隐私保护，在交互和数据分享的各个环节都应用到最新的隐私保护技术。此外，在用户界面的设计上，让隐私相关的说明、采集信息的告知更加醒目、易懂，确保用户理解隐私条款的内容，同时帮助他们了解隐私工程技术能够保护相应敏感信息（Rubinstein and Good，2013）。这两部分同等重要，也已经越来越多地被用于隐私保护实践中。

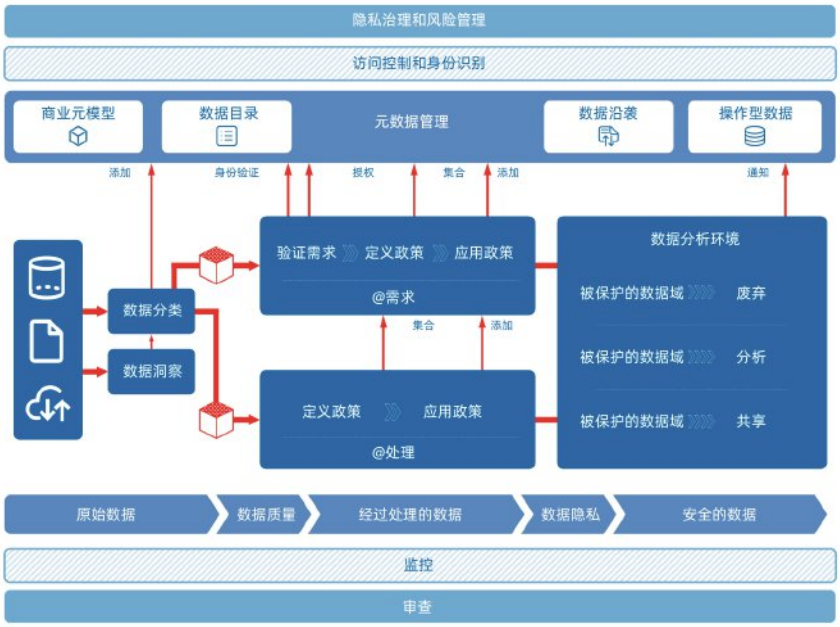


图 25: Privitar 数据处理循环中的数据与隐私保护

资料来源：Privitar.com

注：Privitar 是一个专门为敏感数据提供数据处理的平台。

隐私保护工程化旨在指导数据收集者、处理者和软件开发人员将核心隐私原则转化为具体的设计功能和方法论。在 Gurses 等（2011）研究者工作的基础上，经济学家 Hoepman （2014）确定了在设计软件时纳入隐私因素的八种方式：“最小化、分离、聚合、隐藏、通知、控制、实施和展示。”每个设计策略都可以应用下文介绍的隐私增强技术（Privacy-Enhancing Technology ，简称 PET），开发人员可使用这些技术来实现“隐私设计模式”，并让它们在不同环境下可被复用，以应对隐私相关的设计问题。

无论何种应用中，隐私工程化的基本思路都是将个人数据的收集和处理限制在必要的最低限度。此外，数据生产者需要在收集之前获得用户的授权，在分析和投入使用之

前可以使用假名对数据进行匿名处理。

如今，越来越多的互联网服务都应用了隐私工程化的方法。例如，一个提供隐私保护解决方案的数据平台 Privitar，在所有操作中采用一套以用户为中心的原则（图25）。通过一个三阶段数据隐私流程，使数据生产者能够自动设计数据流，在数据生命周期的全链路实现隐私保护的最佳实践。三阶段包括“原始数据”、“受控数据”和“安全数据”。通过用户授权公司在业务中收集到原始数据，其中的个人信息被脱敏之前，被视为高风险，对原始数据的访问将受到严格控制。原始数据通过数据编目、加密和去标识过程成为“受控”数据。系统会用隐私计算的方法进一步加密，为受控数据创建一个受保护的数据域，最大程度避免数据泄露。在这一安全数据域内，得到授权的分析师可以在域内使用数据，用于特定目的分析工作。通过这种设计，隐私风险可以降至最低。

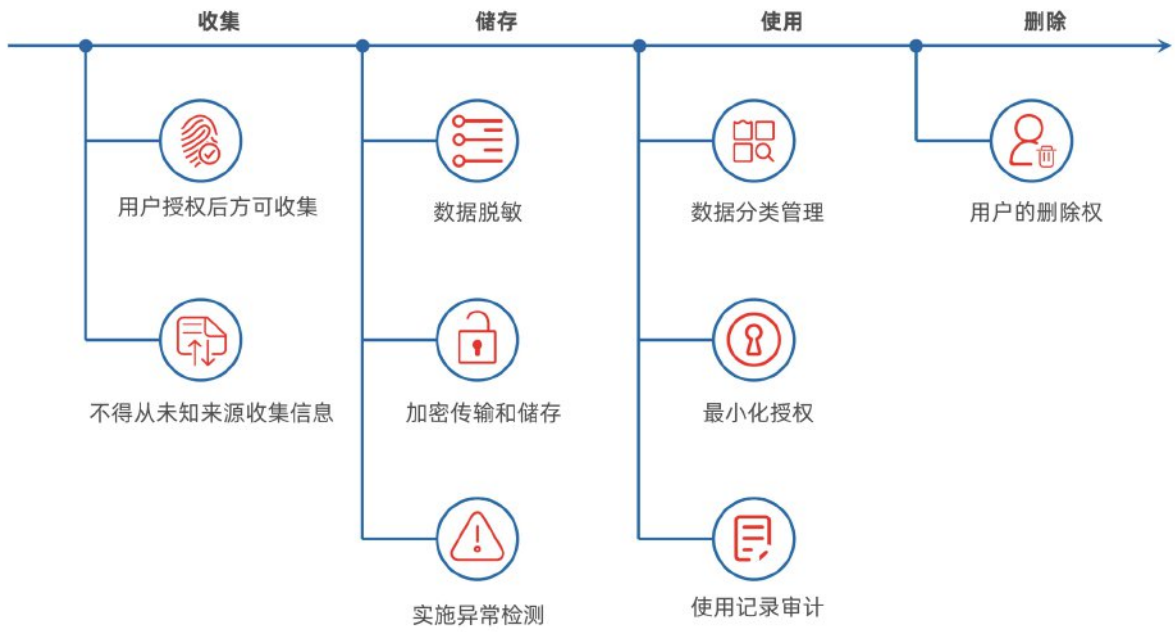


图 26: 隐私保护的机制设计

资料来源：罗汉堂。

在中国，蚂蚁集团已经将以用户为中心的原则应用到其数据使用的整个生命周期中（图26）。在**数据收集**阶段，公司必须获得用户授权，并且确定收集数据的必要性，同时禁止从未知来源收集数据。在**存储阶段**，数据在进入到分析使用环节之前，可以对数据进行脱敏和加密处理，以防在发生数据泄露时，数据中的信息不会泄露。同时，一个实时、全天候的监控系统会自动监测数据分析和使用过程中的异常行为，以最大限度地降低隐私风险。在**使用阶段**，不敏感的加密数据可以在数据分类管理下使用。最终，用户可以选择行使其删除权，抹去记录下来的个人信息。

类似的，苹果公司提出了管理数据生命周期的四个原则：（1）最大限度地减少个人数据收集，（2）在终端设备处理数据以限制不必要数据流动，（3）基于授权管理的透明

性和数据控制，(4) 安全的数据处理流程。这些原则旨在减少数据流处理过程中的隐私风险。苹果公司还使用“差分隐私”技术，给数据集引入噪声，通过放弃部分数据精度来更好地保护隐私。例如，为了在不侵犯用户隐私的情况下弄清用户喜欢什么表情，每次用户点击一个表情，另一个随机表情也会和用过的表情一起被发送到数据集。无须对用户所有的活动进行精准的追踪，数据仍然足以提供有用分析所需的信息。

隐私增强技术（PET）主要针对不可信和潜在有害的数据收集者（Gürses et al., 2011），作为隐私工程化的有效补充。一般把隐私增强技术分为“硬 PET 技术”和“软 PET 技术”。硬 PET 技术利用各种“硬核”技术来降低误判可信第三方的风险。这些技术包括匿名通信渠道（对服务提供商隐藏用户的 IP 地址，同时允许通信），选择性披露凭证（允许用户对自己进行认证，并证明他们有权使用系统，而无须披露其他信息），零知识证明（允许一方向另一方证明一项陈述是真实的，但除了陈述的真实性之外无须透露任何信息），和多方安全计算等（在多方同时计算时通过机密算法只输出结果，并且不可回溯）。

以多方安全计算（multiple-party calculation ，MPC）为例，该技术被广泛用于实现各方共同提供各自的数据，用于彼此的计算分析，同时达到“零知识证明”，即除验证彼此的计算结果外，不提供任何信息（案例 2）。通过该技术，分析师可从多方的数据中获得洞察，而不用接触到各方掌握的“原始”数据，同时原始数据不能通过计算结果进行回溯，化解了各方对数据泄露的担忧。通过这种方式，无须共享原始数据就可以实现多方的数据协作，它可以放大数据的价值，同时大大降低隐私风险。

案例 2 多方安全计算

假设有两位百万富翁 A 和 B，他们都想知道谁更富有。由于他们都担心隐私，不愿意向任何人披露自己的财富。这时，是否有一种方法可以在不侵犯隐私的情况下告诉这两位百万富翁究竟谁更富有呢?

图灵奖获得者姚期智在 1982 年的一篇文章（Yao,1982）中首次提出了这个百万富翁的问题。这篇文章中提出了多方计算的协议。随着大数据、云计算、区块链和隐私意识的发展，近年来采用多方计算变得切实可行。多个采用该技术的系统问世，并投入到商业实践中。

多方计算如何运行呢? 首先，数据所有者对各自的数据进行加密，并将其上传到计算日志，预先设置好的算法在云计算系统中对数据融合分析。授权的数据分析师随后发起查询时，云计算系统在隐私保护情况下处理查询，查询结果将以加密格式传送到授权分析师。

由于未加密状态下数据不出现在源头之外的任何环节，多方计算适合解决数据使用中的隐私担忧。在加密步骤中可以使用各种加密方法，如同态加密、零知识证明、同态承诺等。

通过把数据变得“可用而不可见”，多方计算协议还可以促进数据生产者之间的数据流动。科技公司往往将它们的数据视为专有的业务资产，即使对所有相关方都非常有益也不愿意合作，主要就是担心在此类合作期间造成数据泄露。多方安全计算解决了这一问题，因而有很大的商用价值，近年来发展迅速。

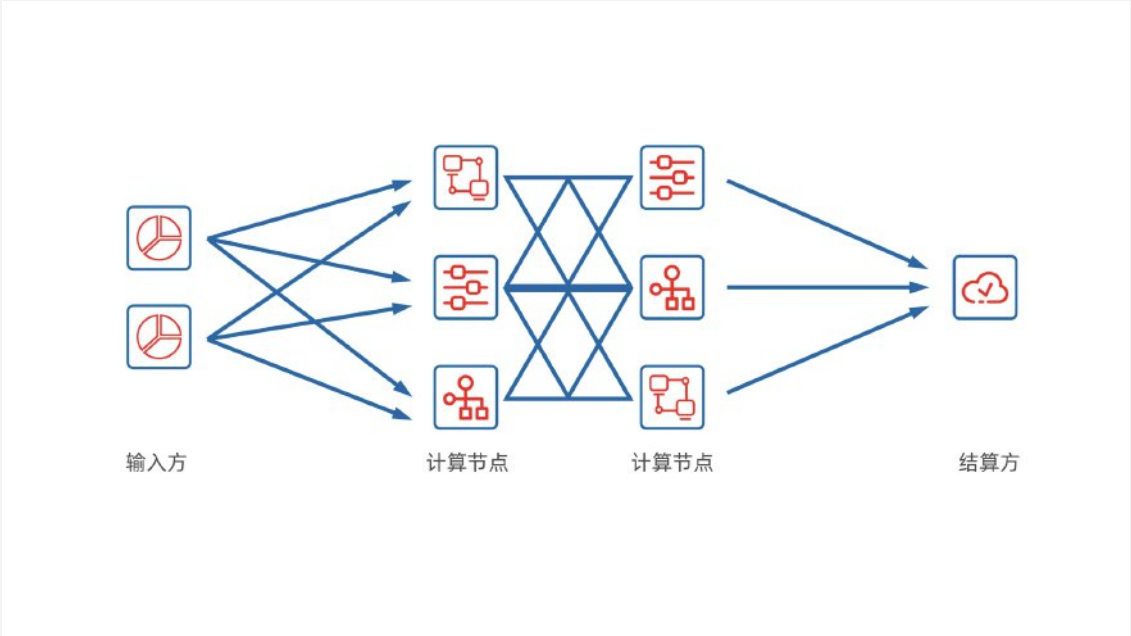


图 27: 多方安全计算

资料来源：罗汉堂。

“软 PET 技术”是一套数据管理工具，可以帮助用户自己做出更好的决策，与服务提供商共享数据，同时满足知情同意的要求，如 cookie 管理工具、隐私仪表板、广告图标等。这些工具背后的核心概念，是用户能够自己决定他们希望对数据收集者的授信程度，因而确保数据在各种环节中的知情权和控制权。

硬 PET 技术成本可能很高（图28）。复杂的分布式计算系统和加密算法要求强大算力。例如多方安全计算系统就对算法的复杂度要求很高，在计算过程中，平台和各方的工程师之间有许多反馈回路，一个简单的结果需要反复的数据请求和计算才能得到，同时针对每种不同的应用场景和计算逻辑，该系统都需要重新定制。这样一个系统的建立和维护都需要大量的资源和人力。

目前，软 PET 技术比硬 PET 技术应用得更广泛。硬 PET 技术不仅昂贵，需要熟悉加密协议并且掌握相关专业知识，还要不断权衡商业应用场景下到底需要哪些数据，以节省成本。迄今为止，大部分中小企业和初创企业依然难以承担硬 PET 技术带来的成本。软 PET 技术的成本低得多，由于它们能让消费者直观感受到隐私体验，能很快提高公司的信誉，同时对数据收集和分析施加更少的限制，不论是对隐私保护还是企业来说都更友好。需要指出的是，先进且高度复杂的硬 PET 技术也发展迅速，更像是“未来的浪潮”。与此同时，越来越多的企业开始在其数字化业务中将隐私设计与隐私加强

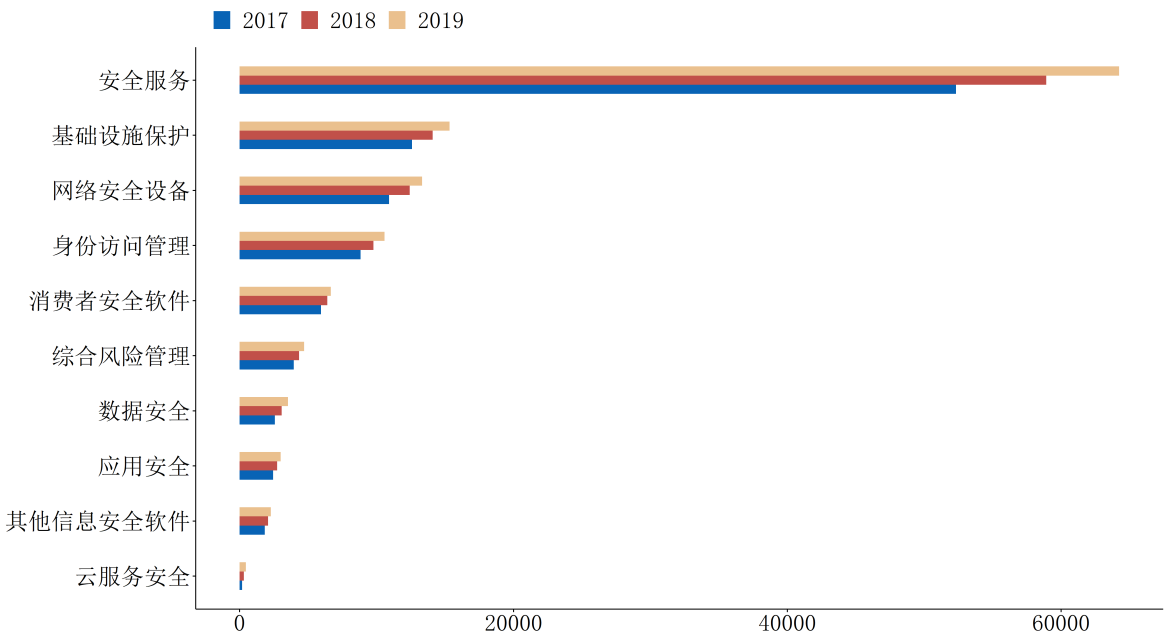


图 28: 2019-2020 全球各项安全类支出预测（百万美元）

资料来源：Gartner2020 年预测。

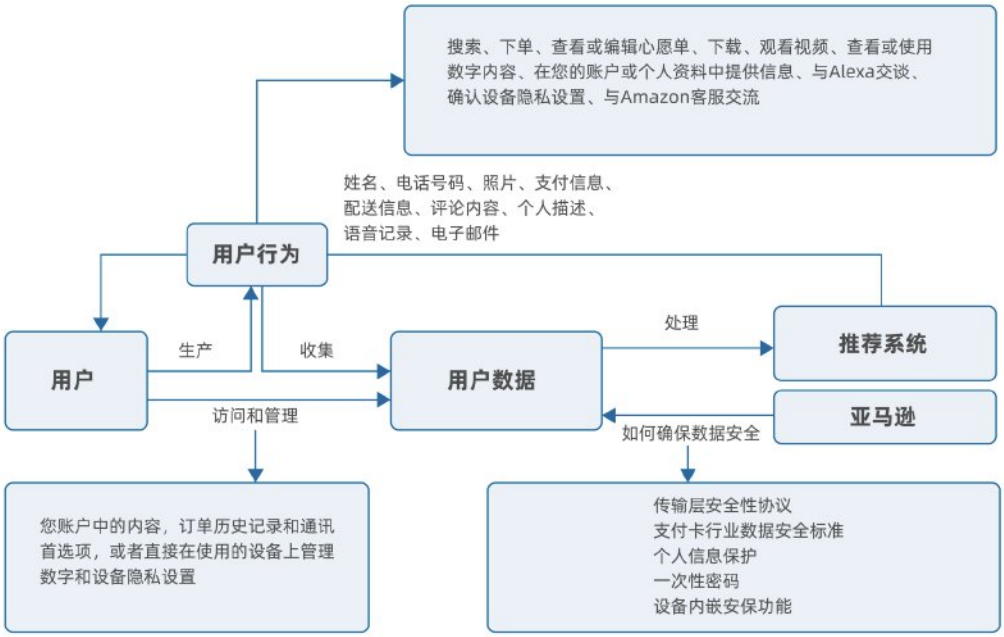


图 29: 亚马逊隐私保护机制

资料来源：亚马逊。

技术相结合，软 PET 技术与硬 PET 技术相结合，提供更全面的隐私保护机制。例如，图29介绍了亚马逊是如何在数据使用中保护数据和隐私。

4.3. 数据安全

数字服务提供商必须保证数据安全。这要求提供商具有较强的内部治理能力，并且能够及时使用最新的安全技术。行业内关于监控和风险管理的“最佳实践”的推广，为科技公司提供了一个指南，使它们能够在整个大数据生命周期中保证数据安全，并且不断更新所用的技术体系。行业自治还包括独立的认证机构、行业行为准则、利益相关者参与公司董事会等[4]。行业内的设计和技术同时提供了前端隐私和下游安全保护，使得企业之间可以分享和输出自己的技术与实践。

图 30: 数据安全管理工具

资料来源：罗汉堂。

加强数据安全性的各类技术也在蓬勃发展（图30）。这类技术和工具包括访问控制平台、数据分类、脱敏工具、审计平台、加密工具等等，可用在数据生命周期的每个阶段。随着数字技术的广泛进步，数据安全技术和工具也在不断升级。例如近十年云计算的兴起，让企业可以在不设置任何离线存储过程的情况下，反而加快数据分析，一方面大大减少了中小企业的数据基础设施成本，另外也提供了更高的数据安全标准，发挥了技术的规模效应。

基于前面提到的数据安全管理工具和数据生命周期管理的理念，构建一个数据安

4 关于此类选项及其潜在有效性的广泛讨论，请参见 OECD（2015）关于行业自律的内容。

全治理框架（图31）非常必要，可确保公司避免不必要的风险，并对意外事件快速做出响应。这样的框架首先需要管理层取得高度一致性，并且得到组织支持，才能畅通运行。《通用数据保护条例》自 2018 年生效后，欧盟和美国公司至少聘请了 2.8 万名数据保护官（DPO）[5] 。某公司建立了一个四层级组织，从战略层到管理层、内部控制层和执行层，用以保护数据安全，数据安全团队占公司总员工人数的 2%。为加强数据保护，还会对所有员工定期举办强制的信息安全意识培训课程。

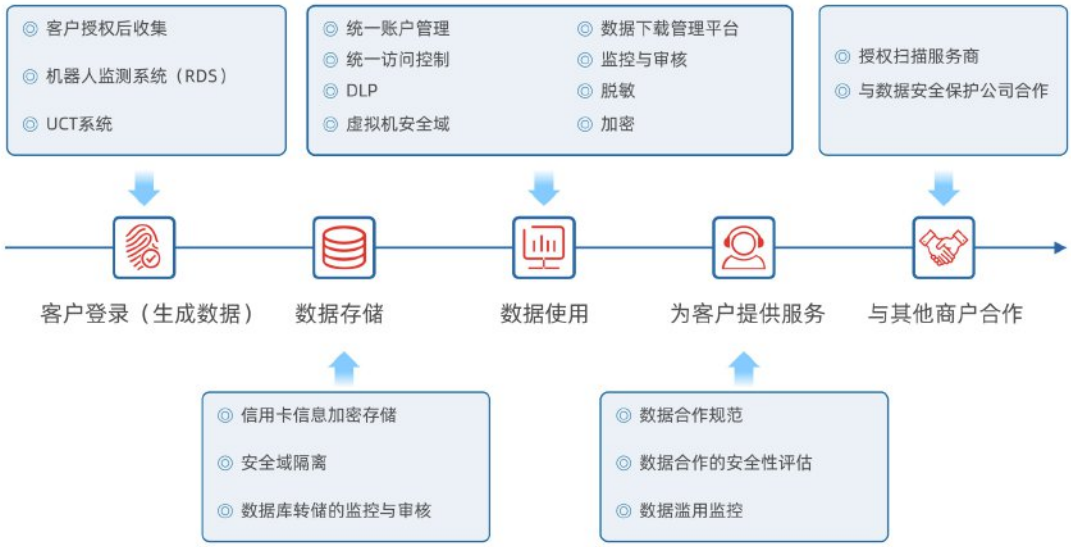


图 31: 数据生命周期中的风险管理

资料来源：罗汉堂。

网络安全本质上是一个攻防游戏。科技公司也会模仿军队的演习手段，创设数据安全的“红蓝军”，通过“战争演习”来测试和改进他们的系统，确保他们能够及时识别并快速应对数据安全和隐私泄露的情况。有的科技公司还会创设出专门的“蓝军”单位，叫做“**网站可靠性工程师**”（SRE，图32），其任务是不断寻找和利用漏洞，定期“攻击”数据和隐私管理系统。这些攻击的目标范围包括数据安全、算法性能、云计算和中台软件等。“演习”也会在各个层面展开，有定期的有随机的，同时会模拟各类极端事件，使得整个系统随时处于预警状态。每年都有一个特殊的月份专门用于确保数据安全，公司内部的任何单位都可能面临来自“蓝军”的随机攻击。这种演习不仅限于对技术基础设施的网络攻击。自 2017 年以来，SRE 团队在其演习中甚至增加了物理威胁，例如模拟自然灾害、断电等冲击，并评估其对平台生存能力的影响。

总体而言，隐私保护和数据安全都需要一个整体的框架来集成技术和面向用户的设计，通过行业基于法规和社会要求的自我治理解决大部分的隐私和数据问题。这越来越

[5] 参见 iapp 报道，At least 28，000 DPOs needed to meet GDPR requirements。

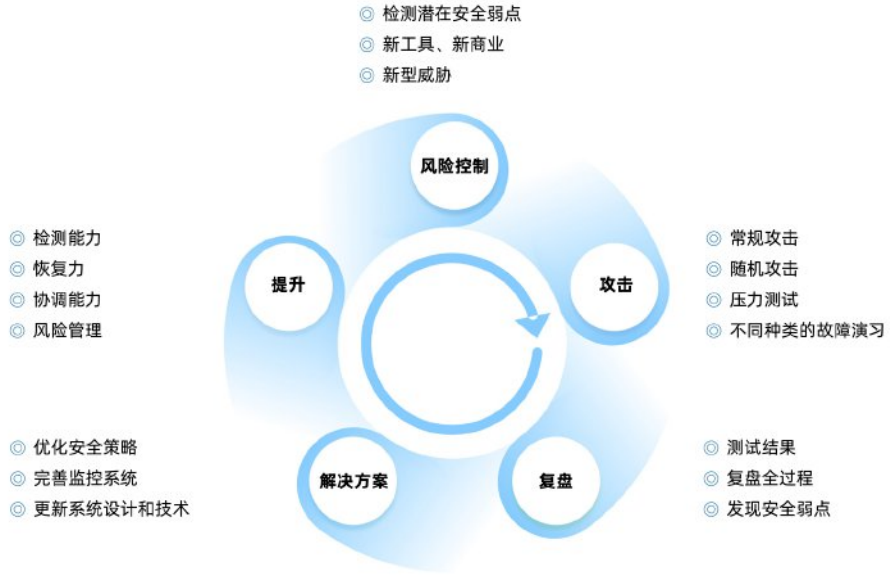


图 32: 某科技公司运用常设的“蓝军”不断提高风控系统

资料来源：罗汉堂。

成为一个社会能够顺利发挥数字革命的价值所必须具备的核心实力。通过机制和技术带来的解决方案，可以显著缓解数据隐私和安全问题。

就像食品行业里的健康和安全一样，当正确的技术应用到位时，现代食品工业中的种类和数量的大爆发不一定意味着更大的安全风险，而是恰恰相反，食物会越来越安全。随着数据隐私和安全日益得到重视，随着时间的推移，更多、更好的技术和机制将变得可用，并成为数字时代许多企业的核心竞争力[6]。我们预计该类技术的成本将迅速下降，促进隐私保护即服务（PPaaS）和数据安全即服务（DSaaS）的发展，将使数以百万的小公司从中获益。重要的是，正如第六章所指出的，对于创新的科技公司和数字服务提供商来说，解决好隐私保护和数据安全这些“痛点”，将受到消费者的青睐和拥抱；当竞争对手的思维还没有转变过来的时候，也是从中争取市场份额的大好机会。

6 监管机构，比如国际清算银行，在法规中开始强调技术的重要性。有关监管和监督科技的广泛讨论，参见 Coeure（2020）。

第 5 章 全面理解数据本质的框架

在上文中我们讨论了数据的价值以及消费者在真实环境下的隐私决策、隐私风险以及应运而生的隐私工程和隐私增强技术。读者不难发现，由于数据具有全新的、与过去有形要素截然不同的本质属性，各利益相关方对数据问题的理解往往局限在主观视角内，容易陷入“盲人摸象”的困局。这背后根本原因在于人们对数据本质的理解不够，或缺乏一个基本共识。因此，要在保护数据安全和隐私的前提下，更好地让数字技术服务于社会和用户，我们需要一个更加整体的和平衡的视角，帮助我们更好地认识数据本质。

在此我们提出一个整体理解数据和隐私本质问题的综合框架——数据权衡框架（图33）。首先，在任何大数据的应用场景中，都包含数据主体、数据生产者，应用场景这三个元素，我们用“数据三角”来概括。另外，数据的两个基本特质，非竞争性和不可分离性，意味着在讨论诸如权属、分享机制、隐私保护等问题时，都不能简单套用传统生产要素的安排方式。基于这些讨论，我们提出了分析数据问题的一个原则，那就是，数据交换是经济活动和创新的基本驱动力，我们需要在促进数据流动的同时，保护数据主体的权利。

图 33: 数据分析框架示意图

资料来源：罗汉堂。

首先是数据三角模型，或者说理解数据的三个视角，能让读者对大数据的本质有更形象的理解（图34）。数据三角包括数据主体、数据生产者以及应用场景。“数据主体”是指数据所描述的各方主体（无论是用于商业还是其他应用）。“数据生产者”是指收集、

处理、存储或分发数据的各方。“应用场景”是指在现实生活中产生、处理和利用数据来促进经济或社会活动开展的场景。

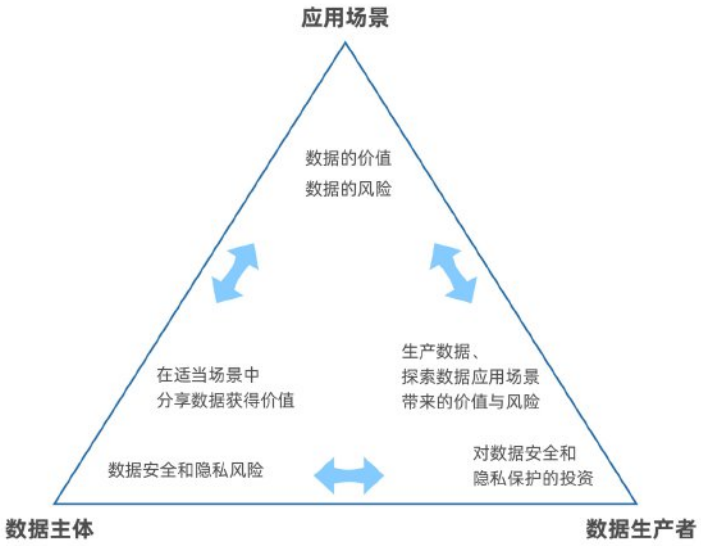


图 34: 数据三角

资料来源：罗汉堂。

其次，数据的两个本质特征决定了数据不同于其他生产要素的权益和责任机制。

第一，数据的非竞争性意味着数据相关主体和数据生产者并非合一的关系，数据可以被多次、多方生产和使用而不消耗数据相关主体。

数据作为一个核心的生产要素，具有非竞争性这一本质特征（Arrow，1969）。这指的是，与石油等实物商品不同，数据可以被无限次生产和使用，而初始数据和数据主体不会被损耗掉。很多人将数据比喻为新的石油，但它更像是可以传递不会消耗的火种。数据分享的范围越广，其价值也会随之增长。诺奖得主 Romer（1990，2018）曾经一针见血地指出，信息是推动人类进步的一种特殊生产资料，但信息的非竞争性这个特点并未得到充分重视。所谓数据生产，我们指的是观察、记录和处理数据。虽然数据主体（个人）可以上报数据，但在大多数情况下，特别是随着数字技术的进步，数据通常由第三方观察或推断得出。举例来说，某人参加会议的事实是由所有参与者，包括其本人的眼睛和耳朵观察而得到的一项数据。

因此，数据不仅是由数据主体生产的，也是由他人生产的。在这个过程中，因为观察角度的不同，每个参与者都可能生产不同的数据内容，或不同版本的数据，这些数据可能在细节上甚至本质上各不相同。另一个例子是消费者搜索和购买过程。数据主体通常无意主动生产数据，也不会在数据生成过程中付出任何直接成本。

如果没有数据主体的活动，当然也不会有数据的产生；但是无论从生产意愿还是生

产成本来说，数据主体都不见得是数据的生产者。数据生产的非竞争性本质决定了，基于对数据主体活动的观察，可以有无数个数据版本。从生产要素的角度来说，把数据的所有权单独归属于数据主体，是极其低效的。数据主体通常不会考虑数据对其他所有数据生产者和使用者可能产生的正外部性，也没有能力聚合和使用数据。非竞争性意味着数据可以有无数个生产者和所有者。将所有权限定给个别所有者，意味着有效信息的供给会严重不足。

其次，数据不等于信息。数据的价值取决于在多大程度上能够提取出解答和服务应用场景问题的信息。这个价值也会因为应用场景不同而不同。

从生成数据到从数据中挖掘出可以回答具体问题的洞见之间，存在着巨大鸿沟。所以在任何机器学习应用中，数据清洗都是重要的第一步。“清洗专家”需要清理数据，因为他们知道自己需要从中提取哪些信息。在观察者眼中，一组数据的价值取决于从中可以获取多少和学到什么样的知识（Blackwell，1953）。

不可忽视的是，将数据处理成有价值信息的成本很高。“信息”不是原始数据。也不是单纯的加减乘除，而是需要训练有素的数据工程师使用科学的方法、流程、算法和系统从许多结构化和非结构化数据中获取知识和洞见。数据科学与数据挖掘、机器学习和大数据等技术息息相关。

数据的价值是在使用场景中实现的。在前面所举的例子中，消费者可以通过对个性化数据的分析获得匹配自己偏好的推荐，小微企业可以通过分析后的数据了解市场。另外一个例子是共享单车服务。想使用自行车的消费者通过扫码进行身份信息登录，可以当场租赁自行车。数据交换是实现这一场景的必要条件。消费者、供应商和提供相关应用的平台，都从中受益。

第三，数据的第二个本质特征是不可分离性（non-separability），即数据使用的效果无法和数据主体完全分离。数据生产者在使用数据时可能会侵犯数据主体的隐私或忽视他们的数据安全，这正是隐私和安全问题的根源。

因此，数据的两个本质特征决定了，一方面，数据主体并非是生产和使用数据的一方，而且往往可以多方同时生产和使用而不会对数据造成损耗，这就和其他物理商品有本质不同；另一方面，数据主体和数据的生产以及使用也无法完全分离：如果数据主体不允许自身的活动被观察到，就不会有数据产生，同时数据的使用也可能会影响到数据主体。**一个合理的数据治理和权益分配机制，应当让各参与方有动机参与到数据的生产、交互和应用，同时保护好数据相关对象的隐私和数据安全**。

第四，虽然数据共享会带来隐私风险，但如果有恰当的机制设计和技术保障（见第四章），风险和收益之间的权衡可以变得可控。

很多先例表明，科技发展带来新的挑战，也带来新的解决方案。比如人们曾经非常担忧乘坐飞机和汽车旅行会有死亡风险，而且这些风险可能永远不会完全杜绝。但有了先进的飞行和汽车安全保障技术，加上政府监管和行业自律，如今很少人会因为在出行和风险之间做出权衡而避免乘坐这两种交通工具。电梯安全问题也与此相似，在纯技术层面，电梯的安全风险几乎可以被完全消除。同样，有了完善法规和先进加工技术，一

个人摄入的食物量与他食物中毒的风险几乎没有关系。数字隐私保护可能永远达不到电梯运行的安全水平，但这不能成为我们停止追求完美的理由。

信息处理和共享曾经是，也将一直是，人类进步的核心基石之一。随着信息共享数量增加到今天的程度，一方面让前所未有的协作成为可能，另一方面其潜在负面影响也在凸显，并受到越来越多的关注。但正如在大多数行业一样，如果有恰当的法规、机制设计和技术，我们就有可能更好地通过分享信息获益，同时将隐私风险降低到更可控的水平，同时享受到数字经济的红利。因此，真正的解决方案不是阻碍数据共享，而是以有效、可持续的方式实现上述目标。

考虑到数据分享对普惠性繁荣的重要意义，隐私风险不应被视为阻碍信息分享的理由。真正的答案不在光谱两端，我们可以运用报告提供的数据权衡框架（图33）来找到中间地带——在促进数据流动的益处与隐私和安全保护的成本之间达成合理的平衡。随着机制设计的改进和相应支持技术的发展，这种权衡本身也在发生变化。在下一章中，我们将讨论到，世界各国有关数据和隐私法规的演进逻辑，是符合这个权衡框架的。

第 6 章 关于数据治理的几个核心问题

随着数字经济发展，全球政府和监管机构对数据治理越来越重视。本章聚焦于数据治理演进的逻辑和方向，并探索大数据对竞争、创新和价格歧视造成的影响，尝试提供相关理论和实证证据。

6.1. 数据隐私

6.1.1 隐私保护原则的发展进化

美国的主要隐私保护监管机构——联邦贸易委员会曾敦促互联网企业“从相关研究中吸取教训，帮助他们最大化大数据带来的益处，同时降低风险”（FTC，2016）。这同样是本章节的目的——尝试理解数据治理法规的逻辑和演化趋势，让企业和政策制定者，在收益和风险之间找到合理的平衡，满足各个利益相关方的需求。我们既不能一味追求利益，忽略隐私风险，也不能矫枉过正，失去获得福利的机会。数字信息革命的到来，在推动数据治理的法规不断演化。不同国家的法规看似不同，但都遵循了一些共通的原则，并且和我们前面介绍的权衡框架的逻辑相吻合。

现代隐私保护法规发轫于《公平信息实践原则》(Fair Information Practice Principles，FIPs）。1973 年，美国健康、教育和福利部（HEW）发布《关于计算机、记录和公民权利》的报告，首次引入了《公平信息实践原则》。该报告呼吁国会出台一个公平信息行为准则，并提出了五大原则[1]。在上述原则的基础上，美国国会通过了 1974 年隐私法案（Hartzog，2016）。FTC 对这五大原则的具体释义为: “(1）通知/知情；(2）选择/许可；(3）接入/参与；(4）完整/安全；(5）执行/纠正”（美国联邦贸易委员会，1998）。不论具体释义如何，FIPs 中提出的这些原则反映了一个基本共识，即数据隐私保护的关键，不是通过对所有权的定义把数据锁起来，而是注重在数据使用过程中的保护。

1980 年和 1981 年，经合组织（OECD）和欧洲委员会分别在《隐私保护和个人数据跨境流动准则》（下为 OECD 准则）和《对个人数据自动处理进行人权保护的公约》[2]中正式采纳了 FIPs，这是它获得国际影响力的标志（OECD，1981）。OECD 和欧洲委员会都明确将个人信息定义为：从收集到储存到传播的每一阶段都需要保护的数据。这两个机构的工作对世界各地相关法律制定产生了深远影响，包括影响力巨大的欧盟《数

[1] 这五项原则包括：(1）不得有任何秘密存在的个人数据记录系统。(2）必须有方法让个人知道记录中有哪些关于自己的信息，以及这些信息是如何被使用的。(3）个人必须有办法防止未经其同意，将为一个目的获得的有关他的信息用于或提供给其他目的。(4）个人必须有方法来纠正或修改关于他的可识别信息的记录。(5）任何创建、维护、使用或传播可识别个人数据记录的组织必须确保数据在其预期用途中的可靠性，并且必须采取预防措施防止数据被滥用。

[2] 欧洲委员会，《对个人数据自动处理进行人权保护的公约》，ETS No. 108 （1981）。

据保护指令原则》[3]、上述 FTC 隐私原则[4]，以及最新颁布的全面隐私法案，如欧盟的《通用数据保护条例》（GDPR）和美国的《加州消费者隐私法案》（CCPA）。

OECD 准则旨在“协调隐私立法，并在维护这一人权的同时……避免国际数据流动中断”。它强调同时做好隐私保护和数据顺畅流动的重要性，这与数据权衡框架的核心原则一致。欧洲发起《数据保护指令》的契机，部分来自《罗马条约》签订后，欧洲国家要建立“共同市场”和“经济与货币联盟”这一雄心勃勃的计划（Cate，2006）。

基于 FIPs，隐私立法的关注点也在随着时间发生改变。早期版本的 FIPs 法案旨在保护个人，免受不公平或虚假信息的侵害，但后来以 FIPs 为基础的法案，特别是自 1980 年 OECD 准则颁布以后，一直以强化消费者对个人信息的控制为目标。最近的隐私保护法案，包括 GDPR 和《加州消费者隐私法案》，进一步加强了消费者的控制权。GDPR 授予了数据主体八项个人数据处理的基本权利[5]。《加州消费者隐私法案》基于消费者权益的五项原则[6]理念起草，其中四项侧重于加强消费者对其信息使用和获取方式的控制权。这些动态、不断改进的原则，是为了通过对数据流动过程中的规定，让个人隐私得到更加有效的保护。

根据美国国际贸易委员会研究，全球数字贸易，包括数据处理和其他数据服务，可以提高生产率和降低贸易成本，进而拉动美国 GDP 显著增长（美国国际贸易委员会，2014）。据统计，数据流在过去十年里带动全球 GDP 增长了约 10 个百分点。因此，GDPR 将个人数据在欧盟范围内的自由流动列为仅次于个人数据保护的重要目标。

为了构建符合数据本质、切实有效的数据治理原则，理解权衡的本质极其重要。正如我们在第四章中所解释，**如果要在数字化信息的利益和成本之间取得平衡，隐私不能被定义为绝对权利，而是应该被视为可选择参与、可以控制被保护程度的权利**。如果一味追求隐私保护，消费者会因为不参与数据交互而舍弃巨大便利。如经济学家 Movius 和 Krup（2009）所言：“保护隐私有明显好处，但也有相关成本。隐私可能带来经济和社会成本；虽然隐私可能保护一些人，但也有代价，比如阻止其他人做出足够明智的决定（Fromholz，2000）。如果将隐私保护视为一个连续体，那么一端是绝对的隐私保护，另一端是完全以经济效率为导向，那么提供隐私保护的成本可被视为放弃一定经济效率和经济安全。因此各国在努力确保公民隐私受到保护时，其实是在两极之间做出权衡。”

要找到“通往中间地带的道路”，“信任”至关重要。美国司法部隐私和公民自由办

3 1990 年，当时的欧洲共同体委员会公布了一项草案：《就个人数据处理和个人数据自由流动保护个人的理事会指令》。

4 从 20 世纪 90 年代中期开始，美国联邦贸易委员会和各州总检察长鼓励美国商业网站运营商采纳并公布在线隐私政策。采取这些政策是自愿行为；遵守这些政策并非自愿行为。该委员会解释了《联邦贸易委员会法》第五节，该节授权联邦贸易委员会起诉“不公平和欺骗性”贸易行为，包括违反既有隐私政策。

5 这些权利包括：知情权、访问权、纠正权、删除/遗忘权、限制处理权、数据可携带权、反对权以及与自动决策和分析有关的权利。

6 这五项基本原则是：（1）信息收集知情权；（2）信息使用知情权；（3）对信息的使用或出售说不的权利；（4）获取信息和请求删除的权利；（5）获得同等服务和价格的权利。

公室首席隐私官兼主任 Peter Winn 指出（Layton ，2019），“信任对于任何机构的效率都至关重要，不管是对公司、国家还是美国司法部本身……人们可能做出错误的选择，在隐私治理中，要么是利维坦式的（绝对权力下的社会控制），要么是完全基于私有产权的自由市场式的。”

6.1.2 隐私保护面临的挑战

即便原则清晰，FIPs 在执行过程中也面临很多现实挑战。不理解这些真实的挑战，隐私保护可能只是纸上谈兵。

FIPs 最初被引入到国家法律时，往往变成流程中的一些简化规定，如获得消费者知情同意的“通知和同意”制度[7] 。信息透明原则简化为“通知”，要求数据主体须了解自己个人信息被使用的内容和方式。“使用限制”简化为过于宽泛且实际上意义很小的“同意”选项，即未经用户同意，为一个目的收集的数据不能用于另一个目的。

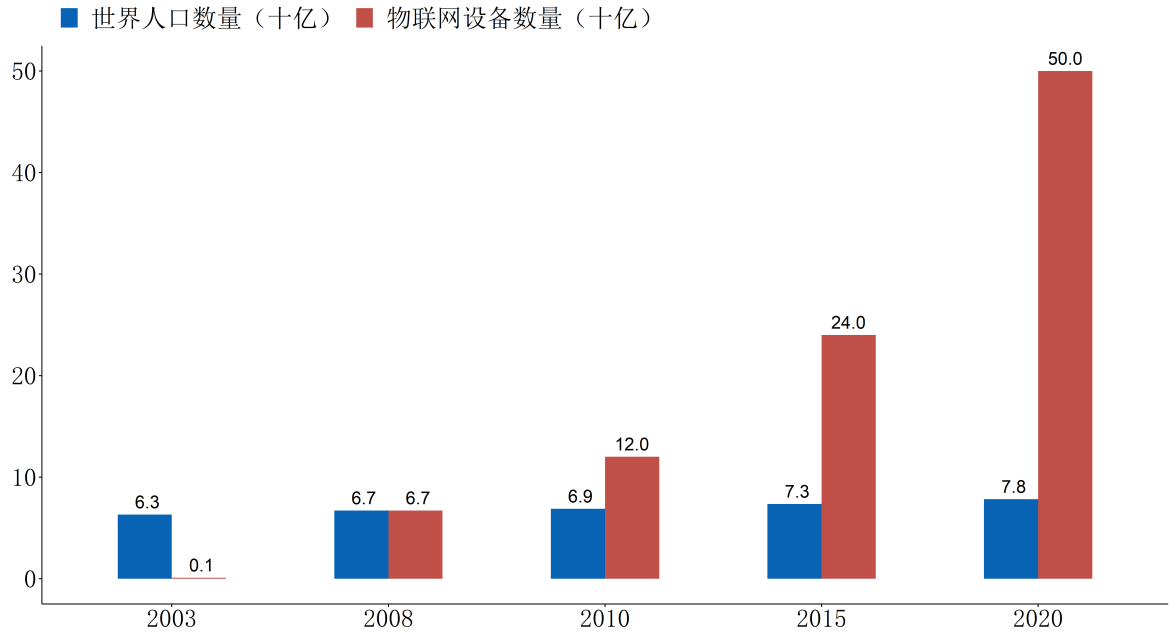


图 35: 物联网爆炸式增长、通知和同意爆炸式增加

资料来源：Cisco 预测，The Internet of Things。

事实上，“通知和同意”制度已经越来越不适合作为现代隐私保护政策的基础。数据量的爆炸性增长（见图35）导致企业更难真正贯彻法规，用户也不得不忍受长篇累牍且难以理解的通知和往往有限的选择。“通知”是 FTC 隐私原则中的“基本核心原则”，但意义不大，面临海量数据时，决策的复杂程度和数量都在上升，获得和给予同意已变得越来越困难。经济学家 Holdren 和 Lander （2014）观察到：“只有在幻想世界中，用

[7] 参见 Cate （2006）。

户才会在点击‘同意’之前真正阅读这些通知并理解其含义。”考虑到物联网的爆炸式增长，即市场上现有数以百万计的互联网连接设备（FTC，2016），阅读通知和给予同意的负担已过大，所以消费者往往无法控制自己的个人数据。如果获得数据使用的许可的成本过高，那么消费者实际上无法许可任何数据使用，因此只会造成我们在第二章开始提到的“隐私悖论”困局。我们需要设计替代方法来应对仍然在飞速增长的数据量。

此外，大数据不仅“大”，还变得日益聪明。大数据可提供超乎想象的丰富信息，但也可能带来对个人隐私想象不到的重大威胁。以当前的通知和同意制度为例。仅从购物历史中追踪维生素购买记录，塔吉特（Target）超市通过大数据就可以推测一个女性是否怀孕[8]，优步（Uber）可以从汽车位置数据中推断出来私人关系[9]。这种滥用数据的行为无法通过简单的通知和同意流程加以防止。“好”的公司可能会意识到潜在威胁，而不会以这种方式滥用数据。“坏”的公司则可能利用这些数据，口头上“遵守”规定，行动上则误导客户或用户。为了保护数据主体不受坏公司的侵害，好公司可能会承受很大的合规成本，这是经济学家 Akerlof 所谓“机会主义行为和道德风险问题”的一个例子。

第三，庞大的大数据网络可能产生严重的负外部性，少数参与者可能以撤回自己对共享关键数据的同意作为威胁，来破坏整个网络的运行，从而引发潜在问题（Landau，2015）。另一种可能是，少数人会愿意披露自己的信息，而这可能会暴露出其他人的相关特征（Hirshleifer，1971）。勾选同意的少数人的选择可能会成为主流，但并未勾选同意的多数人可能会因信息披露而遭受最大损失。在塔吉特（Target）超市的怀孕预测评分案例中，如果有些女性选择分享自己的怀孕状况信息，那么数据分析师可以利用这些人独特的购物特征来推断其他女性是否怀孕，即使后者没有明确同意将预测结果公开。

大数据常常出现不可预测的结果，可能会使 FIPs 中的很多权利变得意义不大。举例来说，考虑到数据的非竞争性本质，以及数据存储的分布式性质，数据生产者实际上不可能定位网络中的“所有个人数据”，更不用说按照数据主体的要求，删除个人数据或在其指定时间授权访问。这种不切实际的要求只会徒增合规成本，以及创造一种掌有控制权的幻觉。

以上论述更证明了发展隐私工程和保护技术的重要性。只有技术不断进步，结合适配的隐私工程，才能更好地应对挑战。

6.2. 数据驱动业务的市场竞争

数字技术正在重塑经济格局，数据驱动的商业行为在竞争中会扮演越来越重要的角色。采用适应这些行为的竞争政策是大数据时代“善治”（good governance）的重要组成部分。其关键是要了解这些新的市场行为中，有哪些可能会促进或阻碍竞争。具体而言，大数据正在多大程度上被用来损害消费者？大数据正在多大程度上妨碍竞争和创新？

竞争或反垄断政策的目标很清晰，就是促进竞争和提高市场效率，从而确保消费者

[8] 参见 Forbes 报道，How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did。

[9] 参见“优步丑闻”网站报道，Analyzed customers’ “Rides of Glory”。

从各方的竞争中获益（Shapiro，2018）。例如，如果数据相关的市场份额或准入壁垒优势是激烈市场竞争的自然结果，那么数据上的“先发优势”会促进产品质量提高和价格降低，使消费者受益；相反，如果该优势实际上减少了市场竞争，阻碍了创新，就会伤害消费者，需要竞争政策的干预。因此，制定合理的竞争政策，需要深入了解商业实践的细节，以及具体分析某个商业行为潜在的危害和好处，同时有能力评估该行为对市场效率的影响，即是否有利于提高生产和销售商品的数量和质量，是否以更低的价格提供商品和服务。

对于和数据相关的所有竞争问题做出整体评估超出了本报告的范畴。以下，我们尝试讨论对几个关键问题的理解。

6.2.1 大数据正在多大程度上被用于损害消费者利益？

理论上，卖家掌握的消费者信息越多，就越有能力向消费者收取差异化的价格，经济学术语称之为价格歧视。在极端情况下，如果一个卖家垄断市场，垄断者能够从消费者身上榨取所有“消费者剩余”（consumer surplus），即向每位消费者收取他们愿意支付的最高价格。而在充分竞争的市场中，来自其他卖家的竞争会极大地限制某个卖家的提价行为，也会限制卖家在不同情况下对同一商品收取不同价格的行为。因此，某个商业行为是否通过阻碍竞争提高了价格，或者得以随意进行价格歧视，赚取垄断利润，常常被当作判断垄断的简化标准。但对于“垄断”的判断，往往不能停留在简化的、“一刀切”的层面，而需要基于具体证据具体分析。

以价格歧视为例，在有些领域中，公司需要价格“歧视”才能生存，也更有助于提高社会福利。这听起来似乎不合逻辑，但公用事业，例如供水、供电等行业，是典型例子。在这类行业中，初始投入巨大，但后续生产的单位成本低（即边际成本低）。一方面，如果市场竞争让价格被拉低至单位成本，那么巨大的初始投入无法被弥补，预见到这种情况，就没有企业愿意投入。另一方面，如果对所有消费者收取同样的价格，那么高额的初始成本会抬高单价，只有极少数消费者可以负担。因此，通过对消费意愿和能力更高的消费者收取高价，价格歧视不仅让提供这类产品的企业得以生存，也同时让更多的人能够参与消费，对整个社会而言，提升了消费者福利。在某种程度上，恰当的价格歧视，事实上起到了促进社会公平、弥合贫富差距的作用。类似的价格歧视现象，在很多其他行业广泛存在，例如，对同样的座位在不同的时间收取不同价格的航空公司，电影院等。

此外，对市场中的不同消费者群体收取不同的价格，有时也可以增进民生福祉，比如剧院、餐馆和其他很多企业对婴幼儿、学生和老年人收取的价格更低，或者制药公司以较低价格向世界贫困地区的公民销售“救命药”。

虽然差异化定价广泛存在，但是一种越来越被社会关注和担心的现象，是基于大数据，对完全相同的产品和服务，只是因为用户身份的不同，就收取不同的价格，这就是所谓的“大数据杀熟”。各国都存在大数据杀熟的个案，引发了社会讨论以及治理机构的关注甚至干预。但一个更重要的问题是：大数据和杀熟的关系有多紧密？在多大程度

上，大数据杀熟正在成为普遍的趋势?

为了回答这个问题，我们首先需要强调，市场竞争的本质，是通过设计、生产出不一样的、更好、性价比更高的差异化产品，因为满足消费者需求而获益。换句话说，价格和产品结合的差异化是最健康的趋势。这也正是约瑟夫·熊彼特称之为的“创造性破坏”（creative destruction）：

> 经济学家总算从唯价格竞争的束缚中挣脱出来。**一旦质量竞争和营销术被容许进入理论圣地，价格变量的统治地位就会不保**……但资本主义的现实景象并非如教科书所描绘的那样，是一种数量竞争。相反，竞争来自新商品、新技术、新的供给来源、新型组织——这种竞争要求的是成本和质量的决定性优势，所打击的并非现有企业的利润率和产量，而是其生存基础。这种竞争的威力更大，更像是炮轰，而非敲门，其重要性足以令常规竞争的效力变得无足轻重；长期提升产量并拉低价格的杠杆总是由其他因素构成（熊彼特，1962）。

在大数据时代，虽然杀熟的个案存在，但是鲜有证据表明，这已经成为任何一个国家的主要趋势。**一个合理的解释是，消费者通过分享大量的个人数据让商家更了解自己，然而这并非一定会带来更多的不利于消费者的价格歧视行为**。这是主要因为商家与消费者的关系正在被数字技术所改变，进而导致了竞争模式的转变。今天的生产者和消费者有着前所未有的直接、高频、长期的连接和互动。普惠性，即以实惠的价格向更多的消费者提供商品和服务，而不是赚尽有限几个消费者的每一分钱，正在成为企业的首要目标（罗汉堂，2019）。最近的研究发现，数字平台对消费者披露而不是隐瞒信息进行价格歧视，会有助于建立信任，长期来讲对平台是最优策略（Ichihashi，2020）。还有证据表明，“网络的透明性限制了实体零售商在不同地区实行价格歧视的能力……这（表明），随着传统零售商与在线零售商的竞争越来越激烈，其地域价格差异将继续减小”（Cavallo，2018，对 Cavallo、Ater 和 Ribgi 的研究总结，2018）。

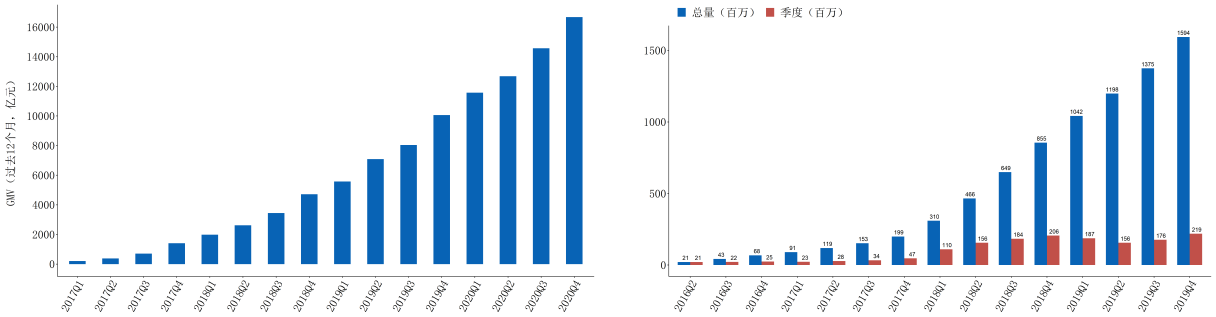
6.2.2 大数据在多大程度上妨碍竞争，进而导致“赢家通吃”的市场结果？

有观点认为，在数据驱动的市场中，由于网络效应（直接或间接）以及随之而来的规模经济，市场壁垒会滚雪球一样越来越高，从而造成了赢家通吃的结果。那么实际情况是怎么样的呢？实证数据表明，至少在中国，与赢家通吃的假设相反，数据驱动型市场事实上竞争激烈，准入壁垒低，呈现高竞争行业的特征。

一是行业头部集中度逐步降低。例如在网络电商行业，尽管阿里巴巴通过不断创新继续增长，但其早期在线上销售领域占据的领先地位，并未阻止新进入者在四处蓬勃增长，在 2015 年至 2019 年的四年时间里，阿里巴巴市场份额下降了 22 个百分点。拼多多销售额在三年内增长了 100 多倍[10]（图36a），吸引了超过 4 亿用户，目前用户规模已经达到最头部电商的同等水平。京东销售额占中国电子商务销售额的 17%，最近成为

[10] 参见文章，Why Can't Taobao Defeat Pinduoduo

“家电市场所有渠道中市场份额最大的平台”。



(a) 拼多多年度交易总额（单位：亿元人民币） (b) 抖音全球首次安装量（全球，单位：百万）

图 36: 新应用可快速崛起

资料来源：感应塔；罗汉堂。
注：（1）抖音于 2016 年面世。2019 年，抖音是下载量第三大的应用（仅次于 WhatsApp 和 Messenger，领先于 Facebook）。（2）抖音的安装量未计入中国和其他地区的第三方安卓下载量。

二是，企业崛起快，衰落也快。Friendster 原本是社交网络行业的“市场领导者”，很快被 MySpace 取代，而 MySpace 在 Facebook 的打压下，几乎已被完全淘汰。2010 年，百度是中国在大数据和人工智能领域的领导者，市值超过腾讯和阿里巴巴，但它现在已经远远落后。抖音的母公司字节跳动则异军突起，用户数迅猛增长，取代百度成为广告收入的市场领导者（图36b）。

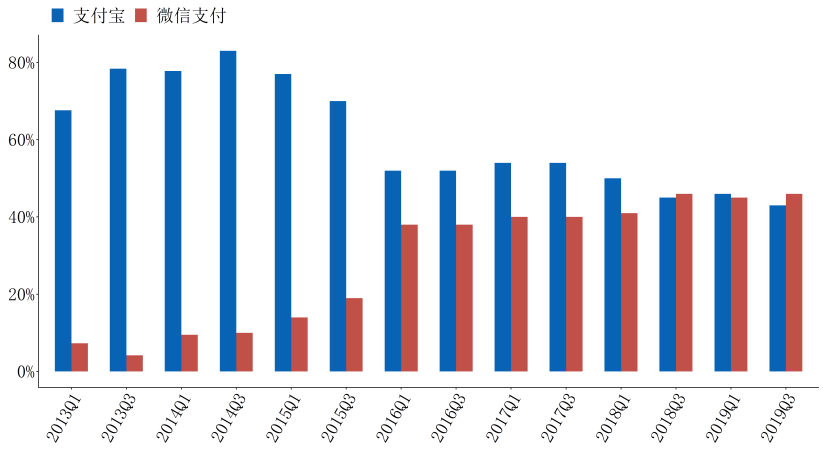


图 37: 中国移动支付市场份额

资料来源：iResearch；罗汉堂

另外一个案例是支付宝和微信支付的竞争（图37）。作为中国数字支付的先行者，支付宝在 2014 年占线上支付总量的近 80%。但到 2019 年，随着微信支付迅速赶上，其

市场份额逐渐降低到 43%。同样，在诸多领域，大数据并没有让早期优势的壁垒越来越高。市场占有率的趋势和赢家通吃的假设背道而驰。

在过去的十年间，标准普尔 500 指数企业的平均寿命呈缩短的趋势，而新入公司的数量却在不断增加，这表明竞争越来越激烈（图38）。随着数字技术的发展，全球商业正进入一个竞争日益激烈的熊彼特世界，一方面是新企业崛起越来越快，一方面是企业寿命越来越短。任何公司想要像过去一样，在某一领域长期保持“高枕无忧”的优势地位，将会越来越难。

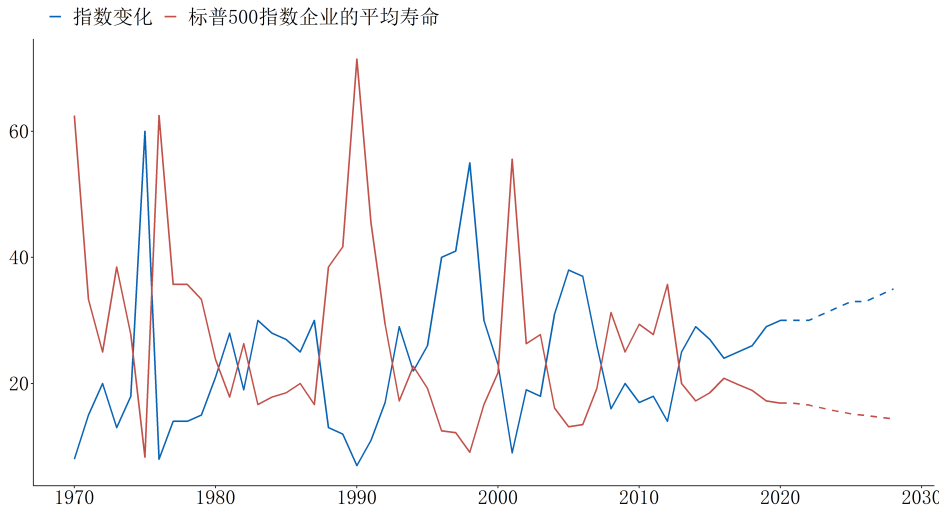


图 38: 标准普尔 500 指数公司的平均寿命

资料来源：标准普尔指数，QAD 博客提供的预测 。

所以至少在大部分领域，大数据并没有造成赢家通吃，大数据可能带来的网络效应和规模效应远不如想象的那么明显。我们认为这是因为以下几个原因：

首先，尽管如第三章提到的，大数据可以从连接、决策、信任三方面提高生产效率，具有明显的商业价值，但大数据只是商业模式的一部分，必然受制于商业模式。例如，虾米音乐最初靠大数据算法推荐得到了许多深度音乐用户的认可，占据用户规模的优势，积累了大量用户数据，但最终由于缺乏盈利的商业模式，在 2021 年关闭。数字经济的发展历史表明，通过移动互联，市场上的新进入者能够以极低的成本互相连接，使得各种细分市场（niche market）频繁、出人意料地出现，技术和新商业模式的结合带来占领市场的机会，然后扩展，威胁到现有公司的市场地位。

第二，数据和有商业价值的信息之间存在巨大的距离，需要数据能力和商业判断力才能发挥数据的价值，存在很大不定性。尽管 Facebook、Instagram 等互联网巨头都拥有广大的用户基础和海量的数据，但仍然难以阻挡抖音在海外市场迅速增长的势头。

第三，互联网用户可以同时使用多个平台，在享受服务的同时在多个平台上分享个人数据。一旦创新者提供新服务满足未被满足的细分需求，就可能成功破局，并迅速积累数据和用户形成正向反馈。拼多多和抖音就是典型的例子。

第四，大数据的有效生命周期很短。新数据源源不断生成，其价值随着时间的推移而下降。相对于现有企业，新进入者不需要创建“相当于现有者规模”的数据存储；相反，他们只需要设计一个策略来积累高度相关和及时的数据（Schepp and Wambach，2016）。

许多研究表明，数据量的优势很少会对竞争对手产生实质性影响。例如，Bajari 等（2019）使用销售数据证明，虽然拥有特定产品相关更多数据有利于更准确的预测，但拥有额外数据的边际价值会下降。随着时间的推移，预测会更为准确，准确性来源于对数据的使用，而不仅仅是拥有更多数据。经济学家 Chiou 和 Tucker（2017）发现，几乎没有证据表明缩短数据的存储时间（在某个案例中，从 13 个月缩短到 3 个月）会显著降低效果。

如经济学家 Lambrecht 和 Tucker（2017）指出，“只有独一无二、稀有、有价值且可持续的资源才能为一家公司提供竞争优势，而数据本质上不具备上述特点。”因此，用经济学家 Shapiro 和 Varian（1998）的话来说，信息产业的市场“主导地位”是脆弱且短暂的：“硬件和软件公司争夺主导地位，因为它们深知目前的领先技术或架构很可能在短时间内被拥有卓越技术的新兴竞争对手推翻。”并不是说大数据的使用绝对不会带来垄断力量，在不同的行业确实存在需要纠正的利用市场地位妨碍竞争的行为，也应当通过各种法规纠正。但是有一点可以肯定，大数据远远不能保证赢家通吃的结果。

6.2.3 公司在多大程度上利用大数据阻碍创新？

由于数据是数字驱动商业模式的一个重要组成部分，善用大数据，无论是通过提供有竞争力的产品吸引客户积累数据，还是通过不断优化算法发挥更大的数据价值，都可以巩固公司当前的地位。如果这种竞争优势是由于具备高效使用数据的能力，就不会妨碍创新，反而会激发创新竞争，并不需要竞争政策的干预。竞争政策的目的不是扶持低效的潜在（或实际）竞争对手。只要大数据本身不阻碍那些高效创新者的进入，就不必对大数据的规模过于忧虑。

相反，如果强迫高效率的公司分享其优势来源，反而“与反垄断法的根本目的有些矛盾，因为这可能会打击（他们及其竞争对手）的积极性……去投资经济上有益的基础设施。强制共享还要求反垄断法院充当中央规划机构，确定准确的价格、数量和其他交易条款，而法院并不擅长于从事这些工作”（Abbot，2018，引用最高法院判决行文，Verizon Wireless 诉 Trinko 案，《美国判例汇编》第 540 卷，第 407-08 页）。

联邦贸易委员会的总法律顾问 Alden Abbot 曾举例说，欧洲和亚洲的某些反垄断机构在较低效的竞争对手的要求下，对领先企业进行“侵入性调查”，为低效竞争对手寻求庇护，从而避免与领先平台竞争。“有益的创新速度将会放缓，影响消费者的福利。更重要的是，由于创新和与市场领导者竞争的动力将会降低，竞争将会减弱。监管机构、公众和政府的偏好将取代可提高消费者福利的商品、服务和平台质量。”他指出，“事实证明这一说法是正确的。尽管欧盟官员多次声明，欧洲政策旨在使欧洲成为数字经济的全球领导者，但所有（西方）大型数字高科技平台公司都是美国公司”（Abbot，2018）。Furman 等（2019）指出，采用大数据技术“可继续以更低的供应商成本、更好的服务、

更好的产品可用性、更好的客户体验为消费者和竞争带来好处……线上平台可成为创新的强大驱动力，它们向消费者提供的服务通常可免费使用”。

从实际效果看，大数据可以从几个方面推动创新。

首先，大数据的 3 个 V 特性成为生产和商业模式创新的强大驱动力；创新优势，而非垄断优势，似乎正在主导数字经济。从教育、商业、医药和金融，到社交媒体、打车、共享单车、观看视频和游戏，几乎在各个数字技术领域发挥深刻影响的行业，其共同点都是数据驱动的创新商业模式，而最具创新性的参与者往往是该行业的新入局者，他们几乎没有初始资本和其他资源，往往在很短的时间内通过创新优势迅速成长。

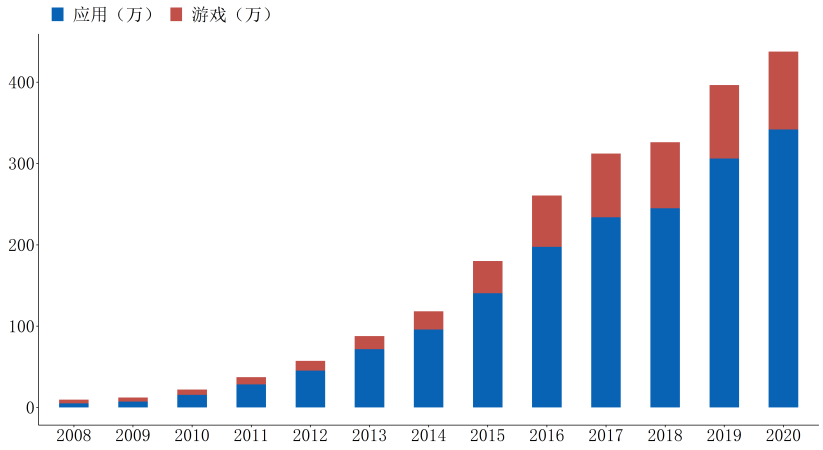


图 39: 2008-2020 年苹果应用商店中可用的应用数量

资料来源：PocketGamer.biz

注：苹果应用商店于 2008 年 7 月 10 日通过 iTunes 更新开放。这恰逢苹果推出支持移动应用的第二代 iPhone 3G。应用商店中的应用程序只适用于 iOS 设备——科技行业一般将苹果的设备生态系统和操作系统称为“带围墙的花园”。

其次，平台作为连接供给和需求的载体，成为创新扩散的重要推动力量。在市场竞争中，平台有意愿用技术改进商业基础设施，推动平台上企业的创新发展。举例来说，自从苹果向第三方开发者开放其应用商店以来，各种创新性的移动应用软件得以迅速发展。根据 Analysis Group（2019）的数据，仅在 2019 年，应用商店生态系统已为全球 5190 亿美元的交易和销售提供了支持。如图39所示，应用商店中可用的应用数量已从最初的 500 个增加到 400 多万个。这些创新应用程序几乎在根本性改变世界娱乐、学习、工作、购物和交往的方式。

苹果的应用商店平台，为移动应用软件开发者创造了一个竞争激烈的市场，促进了手机软件的创新，这也给苹果带来了卓越用户体验的声誉。这种共生关系，一方面通过平台上的竞争促使开发者不断提高其服务质量和竞争力，提供创新的、更好满足用户需求的产品。开发者的成功在很大程度上依赖于用户的评论，从而让满足用户需求与实际的财务激励紧密联系在一起。另一方面，苹果的平台设施（如 TestFlight 和 App Analytics）使开发人员免于承担分销、测试、市场调查等日常繁琐事务，从而得以专注

于自己的核心业务。类似的例子在亚马逊、阿里巴巴、腾讯等平台也大量发生。

除了消费行业，在医疗行业也出现了由大数据支持的促进创新的协作。麦肯锡（2013）通过分析 2011-2012 年“健康数据倡议论坛”参与者的公司和商业模式，发现基于大数据变革的大规模协作创造了大量医疗保健行业的创新。例如，Propeller 是一个基于全球定位系统的记录仪，记录了哮喘患者的吸入器使用情况。通过把即时的信息与已知的引发哮喘的因素（例如，美国东北地区的花粉数量和夏威夷的火山雾）关联，并结合该领域的开创性研究结果，医生可以制定个性化的治疗计划和发现新的预防机会。在这个创新模式中，大数据技术各参与方包括了患者、医生、制药公司、医疗技术公司、医院、医疗保健提供商，他们得以互相协助，发现创新机会，并各自获得收益。

6.2.4 充满潜力的隐私保护市场

随着隐私保护的重要性越来越大，由数字平台自行建立的隐私保护市场正在不断增长，竞争激烈。正如 Entropy Economics 总裁 Bret Swanson 所说，“我们可能低估了公司保护隐私的天然动机。隐私必定会在商业价值主张中占据更大的比例”（Swanson，2019）。

事实上，数字平台之间正在隐私保护领域展开激烈竞争，以寻求更好的方法来提高隐私保护的效率，一方面降低隐私保护的成本，另一方面真正打消用户的隐私保护顾虑，去安全地分享和使用数据。

市场竞争对于隐私保护可能起到的作用至关重要。如果出于对隐私保护的重视，政府对市场施加过多的监管措施，反而可能迫使企业将稀缺的资源，从追求真正有效的隐私保护方式，转移到别处。这样，关于隐私保护的市场竞争就会受到抑制，与最初的政策目标相悖。正如 Swanson 所说，“法律和法规……不能解决所有问题，甚至无法解决大多数问题。不断演化的社会规范、更强大的机构和新的隐私促进技术实际上将承担保护隐私和促进数据流动的大部分重任”（Swanson，2019）。

第 7 章 结语

人类早已意识到，信息的扩散是经济活动的主要推动力和社会繁荣的基础。与以往不同的是，数字化的信息在今天的经济中发挥了革命性作用。在大数据的作用下，生产者与消费者之间的信息联结如此便捷和普遍，让市场的深度和广度出现前所未有的发展。数字技术打破了本地和国际市场的疆界，引入了更强的竞争，生产者和消费者的可能性都得到大幅提升，并在这一过程中，加强了经济的普惠性与可持续性。人类刚刚意识到这种潜力，大数据的价值还有待开发。

历史证明，技术创新催生新的机遇，也会带来新的挑战和质疑。大数据的使用也不例外，人们对此有诸多争论：谁该拥有数据？用户是否从大数据的使用中获得了应有的利益？隐私数据的滥用有多严重？企业和技术能找到保护隐私的解决方案吗？政府又该如何监管数据分享和企业使用数据？

本报告中，我们希望对上述关键问题提供一些新的视角和观点。与传统学术研究相比，我们的核心优势是建立在真实的大数据使用经验上，“用大数据研究大数据”，而非主要建立在理论和数学模型上。正如科斯所说，基于实践，从“竞争的微观分子”中获得真实世界的经验。

要探知用户对隐私保护的真实态度，光靠问卷调查是不够的。更有效的办法，是研究人们在真实环境下的行动和选择，如何权衡数据分享带来的福利与风险。我们发现绝大多数用户愿意通过参与分析个人数据而获得有价值的服务。他们的确在意隐私保护，但个人信息风险应该只是决策的一个维度。用户是否愿意参与分享私人信息，取决于多种因素：服务提供商是否可信，要求数据的敏感度以及服务能带来的价值等。而用户对企业的信任，则取决于企业采取什么样的隐私政策和技术来保护用户隐私。

我们还研究了人们为何愿意参与交换信息，以及信息分享价值的良性循环效应。数据的价值可以总结为三个关键方面：(1) 数据分享带来了前所未有的连接和参与，彻底改变了人类参与协作的范围和深度；(2) 大数据可以带来更明智的决策，让中小企业和贫困人群这些处于信息弱势地位的主体受益；(3) 信息分享在线上的买家和卖家之间建立信任，市场规模因此扩大，引入更多良性竞争。我们的大数据试验表明，一旦关闭个人信息流，产品就无法有效地和消费者进行匹配，整个数字市场规模会大幅缩水，甚至有些市场就此消失。我们的研究证明了，信息交换和数字经济息息相关。让人振奋的是，数字技术的出现为信息集合、扩散和交换等经典经济学问题带来新的解答。

面对隐私和数据安全带来的挑战，我们探索如何通过政府监管和行业自律来降低这些风险。通过合理的机制设计和技术，服务商能在收集和分享数据时，保持匿名性，降低隐私和数据安全风险，让数据自由地流动。随着技术的进步，数据分享与隐私安全不再是不可兼得的“鱼与熊掌”。可以预见，未来人们有可能一方面享受大数据带来的福利，同时保证高效的隐私保护。

数据权属是另一个关键的问题。将数据所有权交给数据相关的主体即用户，看似是

一个自然的选择，但这有悖于数据的非竞争性，让数据使用的效率大打折扣。在现实环境中，个体很少愿意耗时费神地生产和记录数据。经济学家早就指出，让个体提供公共物品通常是低效的。此外，普通的个体并没有建立和挖掘大数据从而促进创新的能力，而数据生产者——科技企业的工程师们拥有这样的能力。

学术界不乏数据所有权归属消费者的支持者，例如在罗汉堂 2019 年会中，一位来自欧洲的学者就提出了这样的观点。然而何志国在 2020 年的研究表明，单一数据所有权会损害所有消费者的利益。将数据的生产者和数据主体分别定义，更符合数据的本质。也许看待数据所有权最好的方式，是将数据主体和数据生产者均视为广义的“数据生产者”，因此也应被赋予隐私和安全的权利和责任。

实际上，数据生产者、使用者和数据主体之间的互动是经济和社会活动的必要条件。数据并没有固定的价值，其价值是在实际的使用场景中实现的。理解了数据生产者、数据主体和使用场景之间的三角关系，就可以消除很多对数据权属和分配机制的误解。在本报告中，我们基于数据的本质特征提出了一个简明的数据分析框架——数据权衡框架，用来理解数据作为生产要素的核心特征、问题和治理逻辑。

对于数据和竞争、创新之间的关系问题，一方面，理论上存在数据被用于阻碍竞争和创新的可能性，现实中也确实存在需要纠偏的案例，但是大数据离赢家通吃的假设有很远距离。政策制定者应尊重不同情况下的事实证据，具体情况具体分析。我们提供了三方面的证据。首先，大数据的兴起催生了以技术和数据为驱动力的全新商业模式；其次，中国市场的很多证据表明，大数据的使用引发了更具活力的竞争局面。鲜有证据表明，企业利用大数据对消费者进行价格歧视或其他行为伤害消费者利益已经成为普遍的现象；相反，企业围绕用户提供各种高性价比的服务，寻求用户的长期认可，越来越成为数字时代的趋势。第三，大数据不仅提升了消费者和供应商之间匹配的颗粒度，也加速了初创企业进入市场和快速发展，促进了市场竞争和创新。

最后我们建议以下数据治理的原则：

原则 1：数据所有权归数据生产者共同享有（包括数据相关主体和其他数据生产者），他们都有保证数据完整性、匿名性以及保护个人隐私的责任。

原则 2：隐私保护和数据安全问题，在很大程度上可以通过把基于法规和原则的隐私保护工程化，并大力发展先进的保护技术解决。

原则 3：在制定竞争和消费者保护政策时，一方面要考虑到特定市场中妨碍竞争和损害消费者的现象，一方面也要认识到大数据对竞争的促进作用和带来的消费者福利。

参考文献

Abbott, A. (2018). Antitrust and the winner-take-all economy. Legal Memorandum, 224.

Acquisti, A., Brandimarte, L., and Loewenstein, G. (2015). Privacy and human behavior in the age of information. Science, 347(6221):509–514.

Acquisti, A., Taylor, C. R., and Wagman, L. (2016). The economics of privacy. Journal of Economic Literature, 54(2):442–492.

Agrawal, A., Gans, J., and Goldfarb, A. (2018). Prediction machines: the simple economics of artificial intelligence. Harvard Business Press.

Akerlof, G. A. (1970). The market for "lemons" : Quality uncertainty and the market mechanism. Quarterly Journal of Economics, 84(3):488–500.

Anwyl, J. (2011). Take polls with a grain of salt. available at https://www.edmunds.com/industry-center/analysis/take-polls-with-a-grain-of-salt.html.

Athey, S. (2017). Beyond prediction: Using big data for policy problems. Science, 355(6324):483–485.

Athey, S., Catalini, C., and Tucker, C. (2017). The digital privacy paradox: Small money, small costs, small talk. National Bureau of Economic Research.

Atkinson, R. D. (2018). How ict can restore lagging european productivity growth.

Bajari, P., Chernozhukov, V., Hortaçsu, A., and Suzuki, J. (2019). The impact of big data on firm performance: An empirical investigation. National Bureau of Economic Research, 109:33–37.

Banisar, D. and Davies, S. (1999). Global trends in privacy protection : An international survey of privacy, data protection, and surveillance laws and developments. The John Marshall Journal of Computer and Information Law, 18(1):1–111.

Berg, T., Burg, V., Gombovi, A., and Puri, M. (2020). On the rise of fintechs –credit scoring using digital footprints. Review of Financial Studies, 33(7):2845–2897.

Blackwell, D. (1953). Equivalent comparisons of experiments. Annals of Mathematical Statistics, 24(2):265–272.

Boisot, M. and Canals, A. (2004). Data, information and knowledge: have we got it right? Journal of Evolutionary Economics, 14(1):43–67.

Carriere-Swallow, Y. and Haksar, V. (2019). The Economics and Implications of Data: An Integrated Perspective.

Cate, F. H. (2006). The failure of fair information practice principles. Social Science Research Network.

Cavallo, A. (2018). More amazon effects: Online competition and pricing behaviors. National Bureau of Economic Research.

Chen, L., Huang, Y., Ouyang, S., and Xiong, W. (2021). Data privacy paradox and digital demands. Working Paper.

Chen, X. and Michael, K. (2012). Privacy issues and solutions in social network sites. IEEE Technology and Society Magazine, 31(4):43–53.

Chiou, L. and Tucker, C. (2017). Search engines and data retention: Implications for privacy and antitrust. National Bureau of Economic Research.

Coase, R. H. (1937). The nature of the firm. Economica, 4(16):386–405.

Coase, R. H. (1994). Essays on Economics and Economists.

Cœuré, B. (2020). Leveraging technology to support supervision: challenges and collaborative solutions. Speech at the Peterson Institute for International Finance, Financial Statement event series.

Culnan, M. J. and Bies, R. J. (2003). Consumer privacy: Balancing economic and justice considerations. Journal of Social Issues, 59(2):323–342.

Dempsey, J. (2019). Institutionalizing the concept of privacy: Global convergence and complexity in the digital age. Speech at the Conference on Privacy and Data Governance organized by Luohan Academy.

Diamond, P. A. (1971). A model of price adjustment. Journal of Economic Theory, 3(2):156–168.

Dinev, T. and Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. Information Systems Research, 17(1):61–80.

Equifax Inc. and Louis Harris and Associates (1996). Equifax-Harris Consumer Privacy Survey.

Federal Trade Commission (1998). Privacy online: A report to congress. available at https://www.ftc.gov/reports/privacy-online-report-congress.

Federal Trade Commission (2016). Big data: A tool for inclusion or exclusion? understanding the issues. FTC Report.

Furman, J., Coyle, D., Fletcher, A., McAuley, D., and Marsden, P. (2019). Unlocking digital competition: Report of the digital competition expert panel. UK government publication, HM Treasury.

Global Privacy Enforcement Network (2018). Gpen sweep 2018: Privacy accountability. Technical report.

Goldfarb, A., Agrawal, A., and Gans, J. (2018). Prediction Machines: The Simple Economics of Artificial Intelligence.

Goldfarb, A. and Tucker, C. (2019). Digital economics. Journal of Economic Literature, 57(1):3–43.

Goldfarb, A. and Tucker, C. E. (2011). Privacy regulation and online advertising. Management Science, 57(1):57–71.

Goldfarb, A. and Tucker, C. E. (2012). Shifts in privacy concerns. The American Economic Review, 102(3):349–353.

Gordon, S. (1989). Darwin and political economy: the connection reconsidered. Journal of the History of Biology, 22(3):437–459.

Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets, volume 70.

Gürses, S., Troncoso, C., and Diaz, C. (2011). Engineering privacy by design. In Conference on Computers, Privacy & Data Protection, Date: 2011/01/25 - 2011/01/28.

Hart, O. D. (1988). Incomplete contracts and the theory of the firm. Journal of Law Economics & Organization, 4(1):119–139.

Hart, O. D. and Moore, J. (1988). Incomplete contracts and renegotiation. Econometrica, 56(4):755–785.

Hartzog, W. (2017). The inadequate, invaluable fair information practices. Maryland Law Review, 76(4):952.

Hau, H., Huang, Y., Shan, H., and Sheng, Z. (2018). Fintech credit, financial inclusion and entrepreneurial growth. Working Paper.

Hayek, F. A. (1945). The use of knowledge in society. The American economic review, 35(4):519–530.

Hirshleifer, J. (1980). Privacy: Its origin, function, and future. The Journal of Legal Studies, 9(4):4.

Hoepman, J.-H. (2014). Privacy design strategies. In Cuppens-Boulahia, N.;Cuppens, F.;Jajodia, S. (ed.), ICT Systems Security and Privacy Protection, pages 446–459.

Holdren, J. P. and Eric S., L. (2014). Big data and privacy: A technological perspective. President's Council of Advisors on Science and Technology.

Hölmstrom, B. (1979). Moral hazard and observability. The Bell journal of economics, pages 74–91.

Hölmstrom, B. (1982). Moral hazard in teams. The Bell Journal of Economics, pages 324–340.

Hölmstrom, B. (2018). Keynote speech at Toulouse School of Economics.

Hoofnagle, C. J., Soltani, A., Good, N., and Wambach, D. J. (2012). Behavioral advertising: The offer you can't refuse. Harv. L. & Pol'y Rev., 6:273.

Ichihashi, S. (2020). Online privacy and information disclosure by consumers. American Economic Review, 110(2):569–95.

Johnson, G. A., Shriver, S. K., and Du, S. (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? Marketing Science, 39(1):33–51.

Jones, C. I. and Tonetti, C. (2020). Nonrivalry and the economics of data. The American Economic Review, 110(9):2819–2858.

Kahneman, D. and Tversky, A. (2000). Choices, Values, and Frames.

Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. Sage.

Kummer, M. E. and Schulte, P. (2019). When private information settles the bill: Money and privacy in google's market for smartphone applications. Management Science, 65(8):3470–3494.

Lambrecht, A. and Tucker, C. E. (2015). Can big data protect a firm from competition. Social Science Research Network.

Landau, S. (2015). Control use of data to protect privacy. Science, 347(6221):504–506.

Laufer, R. S. and Wolfe, M. (1977). Privacy as a concept and a social issue: A multidimensional developmental theory. Journal of Social Issues, 33(3):22–42.

Layton, R. (2019a). Seven deadly sins of the privacy and data protection debate. American Enterprise Institute.

Layton, R. (2019b). Seven virtues of data privacy and protection. American Enterprise Institute.

Layton, R. (2019c). Should online privacy protection be based on trust or control? American Enterprise Institute.

Luohan Academy (2019). Digital Technology and Inclusive Growth.

Martin, K. D. and Murphy, P. E. (2017). The role of data privacy in marketing. Journal of the Academy of Marketing Science, 45(2):135–155.

Maskin, E. S. (2008). Mechanism design: How to implement social goals. The American Economic Review, 98(3):567–576.

McDonald, A. M. and Cranor, L. F. (2010). Beliefs and behaviors: Internet users' understanding of behavioral advertising. Social Science Research Network.

McGann, J. G. (2018). 2018 Global go to think tanks report and policy advice. Think Tanks and Civil Societies Program, University of Pennsylvania Philadelphia.

Myerson, R. B. (1981). Optimal auction design. Mathematics of Operations Research, 6(1):58–73.

Nissenbaum, H. (2009). Privacy in Context: Technology, Policy, and the Integrity of Social Life, volume 58.

North, D. C. and Alt, J. (1990). Institutions, institutional change, and economic performance. Social Science Research Network.

Obermeyer, Z. and Emanuel, E. J. (2016). Predicting the future - big data, machine learning, and clinical medicine. The New England Journal of Medicine, 375(13):1216–1219.

OECD (1981). Guidelines on the protection of privacy and transborder flows of personal data.

OECD (2015). Industry self regulation:role and use in supporting consumer interests.

Pavlou, P. A. (2011). State of the information privacy literature: Where are we now and where should we go? MIS quarterly, pages 977–988.

Phelps, E. S., Alchian, A. A., Holt, C. C., et al. (1970). Microeconomic foundations of employment and inflation theory. WW Norton New York.

Pissarides, C. A. (2000). Equilibrium Unemployment Theory - 2nd Edition.

Pissarides, C. A. (2009). The unemployment volatility puzzle: is wage stickiness the answer? In Econometrica, volume 77, pages 1339–1369.

Reinsel, D., Gantz, J., and Rydning, J. (2018). Data age 2025: the digitization of the world from edge to core. Seagate Data Age.

Romer, P. M. (1990). Endogenous technological change. Journal of political Economy, 98(5, Part 2):S71–S102.

Romer, P. M. (2018). On the possibility of progress. Nobel Prize in Economics documents.

Roth, A. E. (2018). Marketplaces, markets, and market design. The American Economic Review, 108(7):1609–1658.

Rubenstein, I. S. and Good, N. (2013). Privacy by design: A counterfactual analysis of google and facebook privacy incidents. Berkeley Technology Law Journal, 28(2):6.

Schepp, N.-P. and Wambach, A. (2015). On big data and its relevance for market power assessment. Journal of European Competition Law & Practice, 7(2):120–124.

Schumpeter, J. A. (1942). Capitalism, Socialism and Democracy.

Schwartz, P. M. (2003). Property, privacy, and personal data. Harvard Law Review, 117:2056.

Schwartz, P. M. (2004). Property, privacy, and personal data. Harvard Law Review, 117(7):2056.

Shapiro, C. (2018). Antitrust in a time of populism. International Journal of Industrial Organization, 61:714–748.

Shapiro, C. and Varian, H. R. (1999). Information Rules: A Strategic Guide to the Network Economy.

Singh, S. (1999). The Code Book (Vol. 7).

Smith, A. (1776). The wealth of nations.

Smith, H. J., Dinev, T., and Xu, H. (2011). Information privacy research: an interdisciplinary review. Management Information Systems Quarterly, 35(4):989–1016.

Smith, M. D., Bailey, J., and Brynjolfsson, E. (1999). Understanding digital markets: Review and assessment. Social Science Research Network.

Spence, M. (1973). Job market signaling. Quarterly Journal of Economics, 87(3):355–374.

Spence, M. (1974). Competitive and optimal responses to signals: An analysis of efficiency and distribution. Journal of Economic Theory, 7(3):296–332.

Stigler, G. J. (1962). Information in the labor market. Journal of Political Economy, 70:94–105.

Stigler, G. J. (1963). The economics of information. Journal of Political Economy, 69(3):213–215.

Stigler, G. J. (1980). An introduction to privacy in economics and politics. The Journal of Legal Studies, 9(4):2.

Stiglitz, J. E. (1974). Incentives and risk-sharing in sharecropping. The Review of Economic Studies, 41(2):219–255.

Sun, T., Yuan, Z., Li, C., Zhang, K., and Xu, J. (2020). The value of personal data in internet commerce: A high-stake field experiment on data regulation policy. Social Science Research Network.

Tadelis, S. (2002). The market for reputations as an incentive mechanism. Journal of Political Economy, 110(4):854–882.

Tews, S. (2018). Privacy and europe's data protection law: Problems and implications for the u.s. American Enterprise Institute.

UNCTAD (2019). Unctad Stat Data Center.

USITC (2014). Digital Trade in the U.S. and Global Economies.

Veldkamp, L. and Chung, C. (2019). Data and the aggregate economy. preparation for the Journal of Economic Literature.

Verizon (2015). Data Breach Investigations Report.

Volio, F. (1981). Legal personality, privacy, and the family. The International Bill of Rights: The Covenant on Civil and Political Rights, 185.

Warren, S. D. and Brandeis, L. D. (1890). The right to privacy. Harvard Law Review, pages 193–220.

Westin, A. F. (1968). Privacy and freedom. Washington and Lee Law Review, 25(1):166.

World Bank (2016). World Development Report 2016: Digital dividends.

Xu, H., Teo, H.-H., Tan, B., and Agarwal, R. (2009). The role of push-pull technology in privacy calculus: The case of location-based services. The Missouri Review, 26(3):135–174.

Yao, A. C. (1982). Protocols for secure computations. In 23rd Annual Symposium on Foundations of Computer Science (sfcs 1982), pages 160–164.

地址：中国杭州市西湖区西溪路556号Z空间
网址：www.luohanacademy.com

保留特定权利

本报告为罗汉堂研究成果，除非另有说明，其发现、解释和结论仅为署名作者观点，不代表任何罗汉堂的关联机构及其高管和董事 、任何以其他方式参与本报告的顾问和研究者的观点。

就本报告引用、使用或以其他形式包括的任何数据，罗汉堂不保证其准确性。本报告地图和图表中的信息，不代表罗汉堂对任何国家、领域、领土的看法或立场。

报告中的任何内容均不得构成、也不应被视为罗汉堂对任何权利、特权和豁免的限制或者放弃，罗汉堂在此明确声明保留这些权利、特权和豁免。

权利和许可

本报告遵循《知识共享署名许可协议4.0》（CC BY 4.0）（http://creativecommons.org/licenses/by/4.0/）。 根据该许可，在遵守下述条件的前提下，使用者可以复制、发行、传播和改编本报告，包括用于商业目的：

署名 - 请以下述方式引用本报告： 罗汉堂 . 2021.《Understanding Big Data: Data Calculus in the Digital Era 》。 许可证： 知识共享署名许可协议（CC BY 4.0）。

翻译 - 如果您将本报告翻译为其他语言， 请在标明上述署名信息的同时加入以下免责声明： “本译文并非罗汉堂提供， 不应视为罗汉堂的官方译文。 罗汉堂对本译文中的任何内容或错误概不负责。”

改编 - 如果您改编本报告全部或部分内容， 请在标明上述署名信息的同时加入以下免责声明： “本文是对罗汉堂原创作品部分内容的改编。本改编中所表达的观点和意见并未得到罗汉堂的认可， 改编作者对本改编中所表达的观点和意见独立承担全部责任。”

第三方内容 - 罗汉堂不一定拥有本报告所载每项内容的所有权利。因此， 罗汉堂不保证本报告使用的第三方内容不会侵犯他人权利。您应自行承担由此产生的任何侵权风险。如果您希望重复使用本报告中的部分内容， 您有责任确定是否需要对该使用行为获得版权所有者的许可。这里提及的第三方内容包括但不限于： 表格、图表、图像和论述。

如对权利和许可存有任何问题，请询函至浙江省杭州市西湖区西溪路556号Z空间罗汉堂；电子邮件： luohan_service@luohanacademy.com。

数据保护

在研究过程中， 罗汉堂不使用任何个人数据。按照罗汉堂相关政策和流程要求， 因研究所采集和获取的数据均为匿名化数据， 即罗汉堂不会使用任何可识别出特定个人的数据， 也不允许利用这些数据去识别特定个人。罗汉堂出版物中与人相关的数据，都是基于随机样本进行匿名化、汇总后的结果。

罗汉堂严格限制研究数据的使用， 为研究人员和可接触数据的人员制定了数据使用协议， 获得数据访问权限的人员必须接受严格的背景调查， 并签署保密协议， 接受数据安全、 隐私保护方面的培训。经罗汉堂和 / 或关联机构授权， 具有数据访问权限的人员，根据合同或以其他方式，承诺仅将数据用于经批准的研究和内部研讨，承诺不利用数据识别个人身份。

罗汉堂的数据存储在安全的服务器中， 并严格遵循安全流程、 限制访问。数据无法从罗汉堂系统传输到外部服务器或电子邮件地址。罗汉堂采取的数据存储措施遵守严格的数据管理标准。

Luohan Academy

罗汉堂
Luohan Academy

理解大数据：

数字时代的数据和隐私