

企业数据爬取行为的法律风险与规制路径——基于权益均衡的视角

龚鹏程, 张阳阳

河海大学法学院, 江苏 南京 211100

摘要

随着数据需求量越来越大, 为了提高效率, 企业往往采取数据爬取的方式来获取数据。数据爬取引发数据隐私安全、数据污染危及数据治理安全以及不正当竞争破坏市场秩序等多重风险。现有的法律制度存在着爬取行为法律界限模糊、Robots 协议的法律效力未定、数据合规治理机制存在缺陷的困境, 其深层根源在于数据效率与数据安全、数据平台自主经营权与数据接入容忍义务这两对核心权益的失衡。因此, 基于权益均衡的视角, 提出构建多方共治的数据分类分级协同治理体系、完善数据合规治理机制、优化三重授权原则的分类应用、明确 Robots 协议法律效力和保障数据的删除权 5 项措施来破解数据爬取面临的困境, 促进企业的数字化转型。

关键词

数据爬取; 数据治理; 合规; 权益均衡

中图分类号: D912.29

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2026044

Research on legal risks and regulatory paths of enterprise data crawling — from the perspective of balancing rights and interests

Gong Pengcheng, Zhang Yangyang

School of Law, Hohai University, Nanjing 211100, China

Abstract

As the demand for data continues to soar, enterprises often resort to data crawling to enhance efficiency. Data crawling gives rise to multiple risks, including data privacy security, data contamination that endangers data governance security, and unfair competition that undermines market order. The existing legal framework faces challenges such as ambiguous legal boundaries for crawling behaviors, the uncertain legal validity of the Robots Exclusion Protocol, and deficiencies in data compliance governance mechanisms. The underlying causes lie in the imbalance between two pairs of core rights and interests: "data efficiency and data security", as well as "the autonomous management rights of data

platforms and the obligation of tolerance for data access". Therefore, from the perspective of balancing rights and interests, this paper established a multi-party collaborative, classification-based, and tiered data governance system, improved data compliance governance mechanisms, optimized the classified application of the triple authorization principle, clarified the Robots Exclusion Protocol's legal validity as a contract, and safeguarded the right to data deletion. These five measures aim to address the dilemmas faced by data crawling and facilitate enterprises' digital transformation.

Key words
data crawling, data governance, compliance, balance of right and interest

0 引言

数字经济时代，数据已经成为产业发展中新兴的关键生产要素，驱动着生产变革和商业模式上的创新。各行各业积极响应时代号召，推动自身的数字化转型。随着对数据的需求越来越大，为了提升数据获取效率同时降低运营成本，企业通过网络爬虫技术获取外部数据已经成为普遍的现象。虽然爬虫只是一种技术中立工具，但是数据爬取技术在商业应用中的不规范行为产生了诸多风险，包括但不限于数据隐私受到侵害、市场公平竞争秩序遭到破坏以及数据治理出现安全危机。

现有关于数据爬取的法律风险研究主要聚焦在以下3个方面。一是从数据爬取行为本身出发，探讨该技术是否应该得到应用，有的学者从安全角度出发，认为非法爬虫方式侵害了消费者权益，应该禁止：林慰曾^[1]认为，不当爬虫在技术上造成新的数字鸿沟导致在现实中产生新的不平等，通过数据爬取的方式准确地获得了用户画像信息，但是用户却没有获得合理的对价，形成权益不对等局面；有的学者认为在数据爬取问题上的利益权衡应当分场景：丁晓东^[2]认为，应当对平台数据进

行场景化保护，以此平衡企业权益和数据开放问题。二是从不同的实体法角度针对数据爬取带来的问题提出解决之道，江海洋^[3]、侯跃伟^[4]、陈毅坚^[5]等从刑法学角度，周樾平^[6]、陈兵^[7]等从民商法角度对数据爬取行为提出了各自部门法的规制路径。三是对法律规则平衡的研究，聚焦数据效率与数据安全、数据平台自主经营权与数据接入容忍义务的关系。随着技术的不断发展，定量分析方法的应用不断普及，如许可^[8]运用阿列克西的权重公式细化数据权益。

尽管现有的学术研究对企业数据爬取行为带来的法律风险与规制路径提出了宝贵意见，但是仍然存在以下不足：风险的研究呈碎片化，多侧重于对单一风险类型的分析；对法律制度困境产生的原因分析不足，现有的探讨多停留在规范缺失的层面，未能深入揭示缺失规范背后深层次权益冲突的法理本质问题；对策构建的系统性较弱，提出的解决方法偏重某一环节，对于多种措施的协同制度框架缺少研究。鉴于此，本文通过剖析企业数据爬取的风险表征，检视其所面临的法律困境，并从权益均衡的角度探求风险防范的法理进路，构建多层次的法律规制对策，以期为规范数据爬取行为、促进数据要素在市场中合法有序流动提供更具系统性和可操作性的治理参考。

1 数据爬取的风险表征

1.1 数据爬取损害数据隐私安全

数据爬取首要面临的风险是该行为对公民个人信息隐私权构成了直接威胁，其根源在于爬虫技术具有的自动化和规模化的系统特性绕过了个人信息处理时遵循的“告知-同意”核心原则。2019年，中国互联网金融协会发布《关于增强个人信息保护意识依法开展业务的通知》，该文件被视为官方对业内爬虫技术滥用乱象的确认，文件中直接指出一些互联网机构以“大数据”为名，通过“爬虫”业务涉嫌违法违规收集个人信息。实证层面，某些公司因涉嫌非法爬取并交易用户数据而被查处^[9]，揭示了此类行为从个体违规向黑灰产业蔓延的趋势。这些被非法获取的数据流入借贷催收等领域，进而衍生出电话轰炸、曝光通讯录等暴力催收手段^[10]，不仅侵犯了公民个人信息权益，更破坏了金融秩序与社会稳定。种种现象表明，数据爬取已不单单是技术问题，更演变成了一个关乎公民基本权利保障的社会难题。

1.2 数据污染危及数据治理安全

数据污染是指人为篡改或者其他因素造成人工智能算法得到的数据与原有数据不匹配，破坏了原有数据的客观性、真实性^[11]。具体而言，数据污染是指用以开展业务需求的数据从源头或者在交易过程中被恶意篡改并且遭到破坏，导致根据模型得出的结果与真实情况不一致。当需求方根据从外部获得的用户个人数据来对用户进行画像时，由于该部分数据受到污染，

得到的画像并非用户真实反映，阻碍了业务的正常开展，并且清理这部分遭到污染的数据还要付出额外成本。危害性比较强的是数据投毒等针对性攻击，其导致爬取的数据从源头上就已经被污染，若用于训练人工智能模型则直接导致模型偏差进而产生决策失误。还需要考量的是，数据爬取是第三方机构运用技术手段从网页等互联网途径获取用户个人数据的过程，那么在爬取的过程中第三方机构的技术是否得当？又如何确定第三方机构不会篡改爬取到的数据？以及如果爬取的数据本身已经遭到污染，那么爬取之后又将导致二次污染，批量生成的虚假内容影响了用户的选择和喜好等主观领域^[12]。

1.3 不正当竞争破坏市场秩序

不受规则约束的数据爬取行为严重扰乱市场正常竞争秩序，构成不正当竞争。我国司法实践中普遍认可，企业投入大量人力、物力收集加工形成的数据产品享有竞争性权益。淘宝诉美景案中，法院确立了“三重授权原则”，明确未经用户与平台的双重授权，第三方不得爬取和使用非公开数据，这一原则旨在保护数据产品的市场价值，激励企业创新。谷米诉元光案中，法院认为允许经营者肆意爬取他人投入资源形成的核心数据属于不劳而获和“搭便车”行为，实质削弱了原始数据控制者的竞争优势。此类不受约束的数据爬取行为不仅打击了企业数据创新的积极性，长远来看更会导致“劣币驱逐良币”的现象，市场竞争从原有的质量创新竞争异化为数据掠夺能力的竞争，最终损害市场整体效率和消费者安全。所以，对数据爬取行为进行法律上的规制是维护市场健康良好生态的必要之举。

2 企业数据爬取的法律困境

数据作为战略资源在经济发展中的地位日益凸显，随着各行各业对数据的需求日趋增长，不规范的数据爬取行为也带来诸多风险。此类风险在实践当中难以根除，深层原因在于当前法律制度存在以下三重困境。

2.1 爬取行为法律界限模糊

法律的滞后性导致数据爬取行为面临着合法与非法边界模糊的障碍。尽管《中华人民共和国数据安全法》（以下简称《数据安全法》）第三十二条规定任何组织、个人收集数据，应当采取合法、正当的方式，不得窃取或者以其他非法方式获取数据。但是缺乏细化标准，这种立法的缺失直接导致司法裁判陷入困境，法官在裁判时不得不依赖抽象原则进行个案裁量，从而产生裁判标准不一的混乱局面。百度与大众点评案中，法院以百度公司的行为损害了汉涛公司的利益且其行为违反公认的商业道德为由认定百度构成不正当竞争；而在淘宝诉美景案中，法院则创新地提出了“三重授权原则”作为判断标准。这些案例虽然提供了裁判参考，但恰恰反衬出缺乏统一明确的法律规则所带来的裁判结果的不确定性。企业无法对其行为的法律后果形成清晰预期，这就使得企业在日常经营时既可能因为畏惧法律追责而束手束脚，抑制数据正常流通；也可能因为没有明文规定而心存侥幸滋生违法犯罪行为。这种不确定性本身就已经成为制约数据市场发展的制度性障碍。

2.2 Robots 协议的法律效力未定

Robots 协议（又称爬虫协议）是网站所有者通过 robots.txt 文件向爬虫程序表明抓取意愿的技术规范，其核心功能在于沟通协调而非强制阻止。但是这种技术规范的法律效力究竟为何，在立法和司法层面均处于未定状态，效力未定的 Robots 协议构成了对数据爬取监管的又一重困境^[3]。国内外司法裁判对此呈现分歧：国内微梦诉字节跳动案中，司法给出了不同的观点，一审法院认为该协议仅是文字层面的宣示，并非技术措施；二审法院虽对一审结果予以尊重，但也仅将其视为互联网行业应遵守的行业规范，其效力认定需依附于《中华人民共和国反不正当竞争法》（以下简称《反不正当竞争法》）等规定，尤其是在新修订的第十三条（三）中关于数据获取行为的具体规定之下，纵观此案 Robots 协议仅仅是评判主体行为正当性的考量因素之一，而非独立的裁判依据。与之形成鲜明对比的是美国 HiQ 诉领英案，法院甚至作出了更为激进的判决，认为爬取公开数据的行为即便违反 Robots 协议，也如同进入一家未锁门的超市，不构成违法。

这种司法上的分歧揭示了 Robots 协议存在的法律困境：一方面，作为重要的行业自治规范，Robots 协议体现了网站管理者的主观意愿；另一方面，因缺乏法律明文规定，Robots 协议的约束力完全依赖于法官在个案中的自由裁量，在法律上可能被认为不具备法律效力。这种裁判结果的不确定性导致其在实践中陷入无法真正落实的境地，产生的直接后果是数据源网站无法通过法律途径有效预防和制止恶意爬取行为，只能诉诸成本高昂且容易引发技术封堵的手段来维护自身数据安全，这无疑加剧了网络空间的资源消耗，极大地削弱了数据爬虫技术作为数据治理工具的实际效能。

2.3 数据合规治理机制存在缺陷

现有的数据合规治理机制存在系统性缺陷，无法有效应对数据爬取带来的复杂挑战。一是模糊的法律标准导致可操作性不足。虽然《数据安全法》《中华人民共和国个人信息保护法》（以下简称《个人信息保护法》）等确立了数据合规的基本框架，但是原则化、抽象化的规定不利于实践当中的具体操作，《数据安全法》第二十七条要求开展数据处理活动应当依照法律法规的规定，建立健全全流程数据安全管理制度，组织开展数据安全教育培训，采取相应的技术措施和其他必要措施，保障数据安全。利用互联网等信息网络开展数据处理活动，应当在网络安全等级保护制度的基础上，履行上述数据安全保护义务。但对于如何依照缺乏具体的实施细则和案件指引，这种立法上的高度开放导致企业在数据爬取实践中面临着刑事违法性和行政违法性双重认定障碍^[14]，刑事法上可能适用非法获取计算机信息系统数据罪；而行政法上关于过度收集的界限不明，数据合规管理者的具体义务尚未建立，导致企业开展业务时缺乏明确的规则指引。二是企业内生合规动力严重不足。当前我国的合规模式具有强烈的公权力主导色彩^[15]，合规在很大程度上源于规避行政处罚带来的外部压力，并非提升企业自身核心竞争力以获取市场信任的内在需求，这种外生式的合规推动模式，虽然在初期见效快，但是容易导致合规不到位，资源利用效率低下且难以持续^[16]。企业未能将数据合规内化为公司治理的有机组成部分，合规体系缺乏真正的内生活力。三是滞后的监管模式难以适应数字化发展。对于数据爬取这种技术性强、变化比较快的行为，传统的监管模式侧重于事后处罚，迟钝且被动。虽然部分企业建立了内控制度，但是因制

度的效能不同，整体行业缺乏统一的执行标准，无法系统地应对数据爬取带来的全链条风险。

3 数据爬取风险防范的权益均衡进路

前述问题存在的根本原因在于数据效率与数据安全、数据平台自主经营权与数据接入容忍义务之间的权益失衡，为解决这些问题，必须从法理上厘清这些权益背后的均衡本质。

3.1 数据效率与数据安全的均衡

数据爬取规制面临的核心法理困境在于如何平衡数据流通利用的效率价值与数据安全保护的秩序价值。这两大价值并非天然对立，但在资源有限的情况下往往存在优先序上的抉择。

这种价值上的抉择在学术界和司法界引发了讨论，形成了两种鲜明的规制倾向。

一种是以安全优先为导向的权利保护路径，该路径强调数据的人格和安全属性，认为未经授权的数据爬取行为本质上是对个人隐私权、企业财产权和市场竞争秩序的侵害。有学者认为，爬虫技术加剧了数据控制者与用户之间的权力不平等，应对其采用严格规制^[1]。我国司法实践也大多采取此种路径，在谷米诉元光等案中，法院通过《反不正当竞争法》对不劳而获和“搭便车”行为进行惩戒，体现了对既有数据权益和市场竞争秩序的强力维护。另一种则是以效率优先为导向的数据自由流通路径，该路径强调数据的公共产品属性和经济价值，主张对公开数据的获取方式秉持宽容态度，避免形成数据垄断，阻碍创新。美国法院在HiQ诉领英案中的判决即

是此路径的极端体现，将爬取公开数据类比为进入一家未锁门的超市，几乎否定了Robots协议对公开数据的任何限制效力。许可^[6]也认为，在特定条件下，数据爬取具有正当性，其边界需审慎界定。

国内外规制理念的差异折射出对数据本质属性及其法律定位的不同认知偏好。但是任何一种价值的绝对化都有可能带来弊端：绝对安全导向的规制会阻碍数据活力，抑制数字经济发展；而绝对自由导向的规制则会漠视个体权利，侵蚀数据治理的伦理基础。所以，需要在效率与安全动态的均衡中寻找符合实际的举措。数据效率与数据安全之间的关系不是零和博弈，而是相辅相成，安全可信的数据环境是数据要素得以大规模、高效率流通的基础，在降低交易成本的同时促进更广泛的数据协作；高效合规的数据流通机制能为安全技术的研发应用提供更丰富的场景和资源，推动安全治理水平不断提升。对企业而言，数据效率与数据安全的均衡固然需要承担合规成本，但是对二者的投资能够提高企业的长期竞争力。立法者司法者的智慧在于根据数据类型、应用场景、社会公共利益等不同因素，因时因地精细化权衡，在动态发展中寻求特定环境下的最优平衡点。

3.2 数据平台自主经营权与数据接入容忍义务的均衡

数据爬取行为引发的权益冲突在平台层面集中体现为数据平台的自主经营权与其负有的数据接入容忍义务之间的問題。《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》提出的“谁投入、谁贡献、谁受益”原则，为确认平台对其投入资源形成的衍生数据产品享有自主经营权提供了政策依据^[17]。此种权利是激励数据生产和创新的法律工具，但

是数据平台作为数字生态的承载者，也需要承担一定的容忍义务，即在特定情况下允许他人合法接入和利用数据，这种举措是防止数据垄断，促进公平竞争和发挥数据网络正外部性的必要限制。

对数据进行类型化区分以及比例原则的应用为数据平台自主经营权与数据接入容忍义务的均衡提供了缓和的方法。具体来看，首先是对数据进行类型化区分，这是判断平台容忍义务边界的逻辑起点^[6]，可以将平台数据区分为用户生成的原始数据与平台加工衍生的数据产品，前者的数据权益与用户关联更为紧密，平台以自主经营权为由设置普遍访问障碍的正当性较弱，容忍义务范围相对广泛；对于后者，平台投入了实质性劳动和资本，对数据进行了匿名化脱敏、深度分析建模、系统集成等，所享有的自主经营权应该受到更强的保护，平台有权要求他人爬取前需要获取授权，容忍义务范围相对限缩。其次，在可爬取的数据范围内，爬取行为本身也应当受到比例原则的约束以检验其行为的合理性^[18]：目的正当性上，爬取行为须具有合法正当的目的，如提升公共服务、进行学术研究或者提供互补型的创新产品，而非单纯地“搭便车”或进行数据倒卖；手段必要性上，爬取行为应采取实现该正当目的且对平台权益侵害最小的方式，能通过公开API（application programming interface，应用程序接口）获取，就不应采用某些技术手段强行突破反爬虫措施；损害均衡性上，爬取行为对平台造成的潜在损害不应与其追求的利益显失公平。上文的百度与大众点评案中即运用了类似原理，法院认定百度的大量抓取行为超过了必要限度。

数据平台的容忍义务并非无限，其边界取决于数据本身的属性和爬取行为的性

质。对于原始数据，平台需秉持开放精神；对于数据产品，平台可行使自主经营权。但不管在何种情况下，任何爬取行为都必须置于比例原则的审查之下，确保其正当必要且均衡。

4 数据爬取风险防范的法律制度对策

数据爬取提高了企业处理业务的效率，但是不当的爬取模式损害了用户隐私安全，危及数据治理安全，给市场带来威胁，也阻碍了企业的数字化转型。为了解决上述难题，需要在均衡数据效率与数据安全、数据平台自主经营权与数据接入容忍义务的基础上探索新的制度模式，推动企业发展的同时确保数据爬取的合法合规。

4.1 构建多方共治的数据分类分级协同治理体系

4.1.1 规范数据分类分级的监管标准

统一的数据分类分级标准是破解监管执法冲突的前提。当前，不同立法及其配套的法律规范对于数据分类分级的要求存在差异，导致同一数据对象可能同时触发多重且不一致的保护义务。爬虫公司过度抓取用户轨迹数据可能同时面临《中华人民共和国刑法》第二百八十五条第二款的规定处罚、《中华人民共和国民法典》（以下简称《民法典》）人格权编的侵权责任以及《反不正当竞争法》的规制，适用法律复杂且结果不确定^[19]。为终结此种混乱需要超越部门立法的局限，在国家层面进行顶层制度设计，严格落实《数据安全法》第二十一条所确立的国家数据分类分级保护制度，由国家数据主管部门牵头，联合

中华人民共和国国家互联网信息办公室、中华人民共和国工业和信息化部、中华人民共和国公安部等各行业监管机构，吸纳技术专家和法律学者参与，共同制定并发布一份权威的《数据分类分级指南》。该指南应该确立统一的核心数据分类框架，将数据分为公共数据、个人数据、重要数据、核心数据；明确各类型数据的基本定义，在此基础上确定可行的识别标准；划定不同类别数据在爬取、流转、利用各环节的基本安全要求和禁止行为，从根本上杜绝由标准不一导致的法律风险问题。

4.1.2 建立数据爬取黑名单制度

内部可操作的数据爬取黑名单制度是企业将国家宏观统一原则落地为自身微观管理的重要举措，该制度的核心要义在于将《数据安全法》中的核心数据与重要数据以及企业内部认定的敏感数据，明确列为禁止爬取的负面清单。

数据爬取黑名单制度的内容应当包括：一是对数据进行分类，以《数据安全技术数据分类分级规则》等国家标准为根本依据，将企业处理的所有数据标识为核心数据、重要数据与一般数据^[20]；二是明确将核心数据与重要数据自动纳入爬取黑名单，严禁任何形式的爬取行为，绝对禁止爬取可能直接危害国家安全的经济、国防、尖端科技等领域的核心数据；三是对于一般数据中的敏感类别，具体涵盖超过一定规模的个人隐私数据、企业的核心算法参数、未公开的重大经营决策等，企业可以根据自身业务特点评估风险承受能力将其增补进黑名单；四是对于未被纳入黑名单的一般数据及公开数据，其爬取行为也需严格遵循合法正当且必要原则并履行相应的授权程序。

黑名单制度的价值在于通过明确规定

清晰地划出数据爬取的禁区，使数据流通的自由区域更加明确，从而极大降低企业的合规不确定性，确定正面指引和划定负面边界的方式能够增强企业的经营信心。

4.1.3 完善应急处置机制

分类分级制度与黑名单制度的建立明确了数据爬取的规则，而完善的应急处置机制则是应对动态安全漏洞和违规爬取事件的必要补充。企业必须建立起贯穿数据全生命周期的安全监控应急响应流程，对数据进行监测预警，利用技术手段对数据访问，尤其是对爬取行为进行实时日志记录和异常流量监测，设定预警阈值，早期发现单IP地址高频访问、非正常时间访问敏感数据等潜在违规爬取行为。一旦发现数据泄露或恶意爬取事件，应立即启动应急预案，通过立即阻断攻击源、评估事件影响范围、依法向监管部门和受影响用户报告，并及时通过技术手段进行溯源和证据固定，事后应当对事件进行彻底复盘和分析制度漏洞，包括黑名单是否覆盖、权限设置是否过宽等修订完善，实现安全管理能力的全覆盖。

《中国人民银行业务领域数据安全管理办法》等相关法规，正是强调通过有效的管理措施，确保每一位相关业务人员都清晰知晓自身在应急流程中的职责，将分类分级制度落到实处。

4.2 完善数据合规治理机制

4.2.1 细化数据合规律标准

破解数据合规尚未落到实处的困境需要在立法上细化原则性规定，在标准上制定流程化的操作规则。首先，应当充分发挥现有法律框架下国家标准和行业指南的

作用实现监管要求的精细化，《数据安全法》第十七条规定国家推进数据开发利用技术和数据安全标准体系建设。监管机构可以依据《数据分类分级指南》对不同级别的数据进行差异化处理，对于核心数据，严格禁止任何形式的爬取，更拒绝跨境流动，处理权限仅限国家授权机构；对于重要数据，实行事前评估与审批制度，爬取重要数据必须向省级以上网信部门申报数据安全评估，获批后方可进行爬取，爬取过程中需要采取严格加密、脱敏等保护措施；对于一般数据，在遵守《个人信息保护法》中“告知-同意”等基本原则的前提下，允许合规爬取利用。其次，应重视软法的补充作用，中华人民共和国最高人民法院（以下简称最高人民法院）发布典型案例及各地监管机关发布行政执法指导案例，为数据爬取的合规边界提供具有参考价值的行为指引，这些来自实践一线的司法执法案例能有效填补成文法的模糊地带^[20]。最后，企业需要将上述外部要求内化为自身覆盖数据全生命周期的内部操作规程，制定企业内部操作细则，该细则应贯穿数据采集、传输、存储、使用、删除、销毁各个环节：采集环节需进行合法性前置审查，确保数据源正当及爬取手段合法，还需要完整记录数据来源、采集方式、时间等，以满足《数据安全法》第二十七条规定的全流程数据安全管理制度要求；传输、存储环节必须采取加密、访问控制、去标识化等安全技术措施；使用、删除、销毁环节需建立严格的权限管理制度，确保数据在授权范围内使用，在达到保存期限后进行不可逆的彻底销毁。

4.2.2 完善数据爬取合规审查的激励性机制

激励相容的机制能够使合规成为企业

提升市场竞争力的内在需求，该机制需综合运用正向激励和反向约束两种手段。正向激励方面，探索将数据合规投入与税收优惠、专项补贴、绿色通道等政策工具挂钩，对于通过国家认可的合规认证的企业，依法给予一定比例的所得税减免，此举旨在降低企业合规成本，让遵守规则的企业获得实际的利益，引导资源向合规企业倾斜。反向约束方面则严格落实《个人信息保护法》第六十六条等规定，对违法违规的数据爬取行为实施阶梯式惩处，从责令改正、高额罚款，到暂停相关业务或者停业整顿，直至吊销相关业务许可或者吊销营业执照。同时，创新建立企业信用评价机制，由国家市场监督管理总局牵头构建覆盖数据合规水平的企业数据合规信用评价体系，监管机构从纯粹的事后处罚者转变为规则的共同制定者，通过发布合规指南、推行监管沙盒、建立商谈机制等方式来回应技术创新带来的监管挑战^[21]，帮助企业尤其是中小企业理解和落实合规要求，最终形成监管者和被监管者共同推动数据安全流通的良性互动格局。

4.2.3 明确数据平台容忍义务的合规边界

明确数据平台容忍义务的边界，对于划定数据爬取的合法范围至关重要。该边界的确立应综合考量数据属性、平台投入及爬取行为方式三大因素，在平台权益保护与数据流通发展之间找到平衡点。

第一，以数据属性为基本区分标准。对于依法公开的公共数据，数据平台原则上负有较高的容忍义务，无权对合规的爬取行为进行阻止；而对于经过深度清洗、分析、整合的数据报告、用户画像等平台投入实质性资源后形成的衍生数据产品，平台享有受法律保护的竞争性权益，其容

忍义务范围大幅收缩，有权要求他人获取授权。第二，平台引入技术措施作为边界客观化的标志。平台对不愿被爬取的数据采取有效的技术保护措施，如设置Robots协议、登录验证、反爬虫技术等，应被视为其在行使自主经营权、明确拒绝容忍的意思表示，任何绕开或破坏这些技术措施的行为均可初步推定为具有不正当性。第三，爬取行为自身必须符合比例原则。即便是可爬取的数据，爬取行为也须目的正当、手段必要、影响均衡，以研究为目的爬取少量公开评论数据可能具有正当性，但同一主体对相同数据进行全量复制并用于商业竞争，则可能构成权利滥用，超出了平台应容忍的限度。

通过上述数据+技术+行为的三重判断框架，数据平台容忍义务的边界得以清晰可操作，从而为数据平台的合规管理及第三方的合规爬取提供了清晰的法律预期。

4.3 优化三重授权原则的分类应用

为化解传统三重授权原则在大数据场景下高昂成本和流通困难的窘境，本文提出基于数据可识别性（是否可识别特定个人）与平台投入度（是否为平台实质性加工衍生的数据产品）两个关键维度，将数据划分为以下4种类型：第一类数据（可识别性+衍生数据）、第二类数据（可识别性+非衍生数据）、第三类数据（不可识别性+衍生数据）、第四类数据（不可识别性+非衍生数据）。在此基础上，对三重授权原则的具体适用进行精细化分类，构建一个梯度化的授权框架（见表1），以实现数据保护与利用的精准均衡。

该框架的核心在于通过对不同的数据类型进行区分，打破传统“一刀切”的严格授权模式，实现数据流转效率的合理化

表1 三重授权原则的分类适用框架

数据类型	授权要求	法理依据
可识别性+衍生数据	用户同意+平台授权+用户再次授权	保护用户个人信息权与平台的劳动权益
可识别性+非衍生数据	用户同意	数据权益主要归属于用户,平台未付出实质性衍生劳动和资本
不可识别性+衍生数据	平台授权	数据已经匿名化,不涉及个人信息,但是平台投入了加工成本
不可识别性+非衍生数据	遵循必要性原则,可无须授权	数据已经公开且没有特定权益主体,但是爬取行为仍然需要合理适度

提升。

对于第一类数据，因为涉及个人核心信息与平台核心资产，必须坚持最严格的三重授权原则，司法实践对此已在微博诉脉脉案等判例中予以确认。

对于第二类数据，数据权益主要源于用户，平台仅提供托管服务，因此爬取此类数据仅需获得用户同意无须平台的额外授权，避免了平台以自主经营权为名对用户数据的流通进行不合理限制。

对于第三类数据，因其已经不可识别特定个人，不再适用个人信息保护规则，但由于平台投入加工劳动并生成了数据产品，所以爬取行为仍需获得平台授权，以保护其竞争性权益。

对于第四类数据，为已公开的非个人信息，可以纳入合理使用范畴，爬取行为只需遵循《民法典》等法律规定中的必要性原则，避免对数据源网站造成过度负担^[22]。

此分类应用方案并非削弱对数据的保护，而是将有限的合规资源集中于真正高风险、高敏感的数据处理活动上，这是对三重授权原则的具体适用，符合数据要素市场化配置的内在要求。

4.4 明确 Robots 协议法律效力

赋予 Robots 协议明确的法律效力，最直接且可行的路径是在司法实践中将其认

定为介于要约与格式条款之间的默示合同，法理依据在于网站已经通过发布 robots.txt 文件向所有爬虫程序清晰地表达了其关于数据抓取的意愿和规则，这是一种要约。爬虫程序在知晓该协议内容后仍选择访问该网站即可视为以行为默示接受该条款，双方之间成立了一份关于数据抓取规则的合同。这一解释路径完美契合了《民法典》第四百七十一条关于合同订立方式的规定，也为《反不正当竞争法》第二条的适用提供了更为坚实和具体的行为规则基础。由此产生的法律后果是清晰的：善意爬虫方遵守 Robots 协议，意味着其抓取行为获得了网站的默示许可，奠定了其行为合法性的基础；对于违反 Robots 协议者，其行为构成违约，网站可依据《民法典》合同编要求其承担停止侵害、赔偿损失等责任。这种违约责任举证更容易，法律关系更清晰，对于恶意爬虫方，其故意违反协议的行为可直接依据《反不正当竞争法》第十三条第三款等规定认定为不正当竞争行为，需承担相应的法律责任。

在立法暂时缺位的情况下，通过最高人民法院发布典型案例或司法解释，明确 Robots 协议的合同属性，是当前以最小成本激活这一行业规范，大幅提升数据爬取规则确定性的最优方案。

4.5 保障数据的删除权

保障数据删除权是约束数据爬取行为的具体落实，企业不能仅被动地响应用户请求，更应建立主动且可验证的数据删除机制。首先，企业应当将删除权嵌入数据爬取的全流程，在规划爬取项目时，就应在设计文件中预设数据的存储期限和删除触发条件，在项目结束后立即删除或者自爬取之日起满2年自动删除。这不仅是对《个人信息保护法》第十九条的遵守，更是将合规管理前置化的体现。同时，企业应当建立差异化的删除策略，根据数据分类分级结果执行删除，通过爬取行为获得的个人数据应严格履行《个人信息保护法》第四十七条规定的删除义务，并采用技术手段确保数据不可恢复；通过爬取行为获得的非个人数据也应遵循企业的存储期限政策，定期清理非必要保留的数据，降低合规风险。最后，监管部门应当构建可审计的数据删除验证机制，督促企业对数据删除操作进行日志记录并向监管机构提供已履行数据删除责任的证据，这将促使数据删除权从一项抽象的用户权利落地为一项可被监管被审计的企业内部管理流程，真正实现彻底的删除。

5 结束语

本文系统研究了企业数据爬取行为的风险、困境与规制路径，从权益均衡这一法理学核心视角切入，在理论层面突破了以往就事论事的分析框架，为理解数据爬取规制难题提供了更具解释力的理论工具；在实践层面提出的数据分类分级+爬取黑名单治理框架、对三重授权原则基于数据类型的精细化分类应用，以及将Robots协议明确为具有合同法律效力的建议为立法、

司法和企业合规提供了清晰且可操作的方案设计。本研究也存在一定的局限性：一是主要采用定性研究和案例分析的方法，缺乏对大样本数据的定量分析，未来可通过问卷、访谈等方式对数据爬取规模、成本收益及企业合规现状进行实证考察，增强研究结论说服力；二是仅立足国内法律制度，未充分探讨跨境数据爬取带来的国际法律冲突问题，这将是未来研究的重要方向。规范数据爬取行为并非扼杀技术，而是为了引导其在合规轨道上发挥更大价值，只有在法治框架下平衡各方权益，才能真正释放数据要素潜力，护航数字经济行稳致远。

参考文献：

[1] 林慰曾. 数据爬虫技术对金融信息安全的冲击及制度回应[J]. 北京航空航天大学学报(社会科学版), 2022, 35(4): 161-169.
Lin W Z. Impact of data crawler technology on financial information security and the institutional responses[J]. Journal of Beijing University of Aeronautics and Astronautics (Social Sciences Edition), 2022, 35(4): 161-169.

[2] 丁晓东. 数据到底属于谁——从网络爬虫看平台数据权属与数据保护[J]. 华东政法大学学报, 2019, 22(5): 69-83.
Ding X D. Who owns the data platform data ownership and protection from the perspective of web crawler[J]. ECUPL Journal, 2019, 22(5): 69-83.

[3] 江海洋. 数字时代数据爬取的刑法规制：法益界定与数据确权[J]. 比较法研究, 2024(2): 149-163.
Jiang H Y. Criminalization of data scraping in the digital era: definition of legal benefit and data entitlement[J]. Journal of Comparative Law, 2024(2):

149-163.

[4] 侯跃伟. 共享视角下数据爬取行为刑法规制理念重塑与路径展开[J]. 江苏社会科学, 2024(2): 165-174.

Hou Y W. Reconstruction of criminal law regulation concept and path expansion of data crawling behavior from the perspective of sharing[J]. Jiangsu Social Sciences, 2024(2): 165-174.

[5] 陈毅坚, 曾宪哲. 网络爬虫刑法规制研究[J]. 广东社会科学, 2022(5): 240-253.

Chen Y J, Zeng X Z. Research on criminal law regulation of web crawler[J]. Social Sciences in Guangdong, 2022(5): 240-253.

[6] 周樾平. 数据爬取的不正当竞争认定规则研究[J]. 南大法学, 2023(2): 87-102.

Zhou X P. Research on rules of law on unfair competition practices for data crawling[J]. Nanjing University Law Journal, 2023(2): 87-102.

[7] 陈兵. 保护与竞争: 治理数据爬取行为的竞争法功能实现[J]. 政法论坛, 2021, 39(6): 18-28.

Chen B. Protection and competition: realization of the function of competition law in governing data crawling[J]. Tribune of Political Science and Law, 2021, 39(6): 18-28.

[8] 许可. 数据爬取的正当性及其边界[J]. 中国法学, 2021(2): 166-188.

Xu K. The legitimacy and the boundary of data scraping[J]. China Legal Science, 2021(2): 166-188.

[9] 高永明, 马光远. 网络爬虫的规制逻辑: 行政监管前置的路径提倡[J]. 吉林大学社会科学学报, 2024, 64(3): 73-86+236.

Gao Y M, Ma G Y. The logic of regulating web crawlers: advocating for preceding administrative oversight[J]. Jilin University Journal Social Sciences Edition, 2024, 64(3): 73-86+236.

[10] 山西经济日报. 消费金融暴力催收乱象多: 电话打爆通讯录 逾期还贴大字报[EB]. 2024.

Shanxi Economic Daily. Rampant malpractice in consumer finance: aggressive debt collection tactics exposed—from bombarding contact lists to public shaming posters[EB]. 2024.

[11] 张喆锐. 数据污染的平台治理责任[J]. 东方法学, 2026(2): 100-115.

Zhang Z R. Platform governance responsibility for data pollution[J]. Oriental Law, 2026(2): 100-115.

[12] 刘晓春. 数据功能类型视角下数据污染的治理维度[J]. 人民司法, 2024(10): 15-20.

Liu X C. The governance dimension of data contamination from the perspective of data function types[J]. People's Jurisprudence, 2024(10): 15-20.

[13] 刘宪权. 非法获取公开数据行为的刑法规制[J]. 法律适用, 2025, (8): 97-113.

Liu X Q. Criminal regulation on crimes of illegally obtaining public data[J]. Journal of Law Application, 2025, (8): 97-113.

[14] 齐英程. 数据合规协同激励体系的构建与完善[J]. 东方法学, 2024(2): 98-108.

Qi Y C. Construction and improvement of collaborative incentive systems for data compliance[J]. Oriental Law, 2024(2): 98-108.

[15] 汪青松, 邱欢. 合规制度发展的中国范式及其与商法关系探析[J]. 重庆社会科学, 2023(12): 185-202.

Wang Q S, Qiu H. An exploration of the Chinese paradigm of compliance system development and its relationship with commercial law[J]. Chongqing Social Sciences, 2023(12): 185-202.

[16] 刘盛. 现代金融体系视野下的金融法: 理念信守与制度表达[J]. 政治与法律, 2022(11): 80-95.

Liu S. Financial law from the perspective of modern financial system: adhered-to ideas and institutional expression[J]. Political Science and Law,

2022(11): 80-95.

[17] 黄丽华, 杜万里, 吴蔽余. 基于数据要素流通价值链的数据产权结构性分置[J]. 大数据, 2023, 9(2): 5-15.

Huang L H, Du W L, Bi Y W. Structural separation of data property rights based on data factor circulation value chain[J]. Big Data Research, 2023, 9(2): 5-15.

[18] 陈兵, 姚俊羽. 公开数据保护的理念澄清与路径选择[J]. 中国特色社会主义研究, 2024(2): 38-52.

Chen B, Yao J Y. Concept clarification of public data protection and its path choice[J]. Studies on Socialism with Chinese Characteristics, 2024(2): 38-52.

[19] 陈咏梅, 张姣. 跨境数据流动国际规制新发展: 困境与前路[J]. 上海对外经贸大学学报, 2017, 24(6): 37-52.

Chen Y M, Zhang J. New development of international regulation-formulation in cross-border data flow: dilemma and the way forward[J]. Journal of Shanghai University of International Business and Economics, 2017, 24(6): 37-52.

[20] 鄢浩宇. 企业数据合规的困境纾解与体系构建[J]. 华中科技大学学报(社会科学版), 2024, 38(4): 36-45, 71.

Yan H Y. Dilemma relief and system construction for corporate data compliance[J]. Journal of Huazhong University of Science and Technology (Social Science Edition), 2024, 38(4): 36-45, 71.

[21] 薛澜, 赵静. 走向敏捷治理: 新兴产业发展与监管模式探究[J]. 中国行政管理, 2019, 35(8): 28-34.

Xue L, Zhao J. Toward agile governance: the pattern of emerging industry development and regulation[J]. Chinese Public Administration, 2019, 35(8): 28-34.

[22] 徐伟. 企业数据获取“三重授权原则”反思及类型化构建[J]. 交大法学, 2019(4): 20-39.

Xu W. Reflections on the “triple authorization principle” of enterprise data acquisition and its typological construction[J]. SJTU Law Review, 2019(4): 20-39.

作者简介



龚鹏程（1974—），男，博士，河海大学法学院副教授，主要研究方向为商法基础理论、经济法、金融法、证券法、资本市场法律制度。



张阳阳（2001—），女，河海大学法学院硕士生，主要研究方向为经济法、数据法。

录用日期: 2025-09-16
通信作者: 龚鹏程, pcgong@foxmail.com