

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 210003

摘要

随着对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两对核心权益的失衡, 因此, 基于权益均衡的视角, 提出需要建构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 协议作为合同法律效力和保障数据的删除权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 Corporate Data Crawling: From the Perspective of Balancing Rights and Interests

Gong PengCheng, Zhang Yangyang

School of Law, Hohai University, Nanjing 210003, China

Abstract

As the demand for data continues to soar, enterprises often resort to data crawling to enhance efficiency. Data scraping gives rise to multiple risks, including data privacy and security issues, 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 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, it is proposed to establish a multi-party collaborative, classification-based, and tiered data governance system, improve data compliance governance mechanisms, optimize the classified application of the triple authorization principle, clarify the Robots protocol's legal validity as a contract, and safeguard the right to data deletion. These measures aim to address the dilemmas faced by data crawling and facilitate enterprises' digital transformation.

Key words

Data crawling, data governance, compliance, balance of rights and interests

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，驱动着生产变革与商业模式上的创新。各行各业积极响应时代号召推动自身的数字化转型，随着对数据需求越来越大，为了提升数据获取效率同时也降低运营成本，企业通过网络爬虫技术的方式来获取外部数据已经成为普遍的现象。但是爬虫只是作为一种技术中立工具，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了侵害、市场公平竞争秩序遭到破坏以及数据治理出现安全危机。

现有的关于数据爬取的法律风险研究主要聚焦在以下三方面：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，在现实中造成新的不平等，通过数据爬取的方式准确地获得了用户画像信息，但是用户却没有因此获得合理的对价，造成了不平等、不公正。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，平衡企业权益与数据开放；

二是从不同的实体法角度针对数据爬取带来的问题基于解决之道，江海洋^[3]、侯跃伟^[4]、刘营^[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，聚焦效率与安全、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，如许可运用阿历克西权重公式细化数据权益^[8]。

现有的学术研究为企业数据爬取行为带来的法律风险的规制提出了宝贵意见，但是仍然存在以下不足：一是对于风险的研究呈碎片化现象，多侧重对单一风险类型的分析。二是对产生该困境原因的分析不足，法律规制困境的探讨多停留在规范缺失的表面，却未能深入揭示缺失的规范背后数据效率与安全、数据平台自主经营权与数据接入容忍义务等深层次权益冲突的法理本质问题。三是对策构建的系统性较弱，提出的解决方法偏重于某一环节，对于多种措施的协同制度框架缺少研究。鉴于此，本文旨在通过系统的文献梳理厘清现有研究的贡献与不足，在此基础上剖析企业数据爬取的风险表征，检视其所面临的法律困境，并创新性地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操作性的治理参考。

1 数据爬取的风险表征

1.1 数据抓取损害数据隐私安全

数据爬取的首要面临的风险是该行为对公民个人信息与隐私权构成了直接威胁，该风险的根源在于爬虫技术具有的自动化、规模化的特性系统绕过了个人信息处理时遵循的“知情-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》，该文件可以被视为官方对业内爬虫技术滥用乱象的确认与回应，文件当中直接指出部分机构以大数据的名义违规使用爬虫技术收集个人信息。在实证层面，诸如公信宝、魔蝎科技等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向。当这些被非法获取的数据流入到借贷催收等领域，进而衍生出的电话轰炸、曝光通讯录等暴力催收手段^[10]，这不仅侵犯了公民个人信息权益，更破坏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，而是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

数据污染是指由于人为篡改或者其他因素导致通过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，是指用以开展业务需求的数据从源头或者交易过程中被恶意篡改遭到破坏，导致根据模型得出的结果与真实情况不一致。当需求方从外部获得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，阻碍了正常业务的开展，由于清理这部分遭到污染的数据，还要额外付出成本。危害性比较强的是对于数据投毒等针对性攻击导致爬取的数据从源头上就已经被污染，若将其用于训练人工智能模型则直接导致模型偏差与决策失误。还需要考量的是，数据爬取的过程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，那么在爬取的过程中第三方机构的技术是否得当，爬取而来的数据是由第三方获得，又如何确定不会遭到篡改？以及如果爬取的数据本身已经遭到污染，那么爬取之后又导致二次污染。由于批量生成的虚假内容影响了用户的阅读、选择和喜好等认知领域^[12]，数据投毒导致爬取的数据从源头上就已经受到污染，在爬取的过程中又受到人为篡改，极大地破坏了数据的真实性，不利于正常业务的开展。

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，在我国司法实践中普遍认可企业对于其投入大量的人力物力收集、加工而成的数据产品享有竞争性权益。淘宝诉美景案中^[13]，法院确立了三重授权原则，明确了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与激励企业创新。谷米诉元光案^[14]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，使得市场竞争从原有的质量与创新竞争异化为数据掠夺能力的竞争，最终损害整体市场效率和消费者安全，所以对数据爬取进行竞争法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检视

数据作为战略资源在经济发展中的地位日益突显，随着各行各业对数据的需求日趋增长，与此同时，不规范的数据爬取行为也带来诸多风险，上述数据爬取所带来的各类风险之所以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

合法与非法的边界模糊是数据爬取行为面临首要障碍，这种模糊的根源于立法原则性规定的滞后性。尽管《数据安全法》第32条规定数据收集应该合法、正当，但是正当为何缺乏细化标准，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，这就产生裁判标准不一的混乱局面。百度与大众点评案^[15]中，法院以超过必要限度“违反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，但是也恰恰反衬出由于缺乏统一、明确的法律规则而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稳定预期，这就使得企业在日常经营时既可能因为畏惧法律而束手束脚，抑制了数据正常流通，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与协调而非强制阻止，但是这种技术规范的法律效力究竟为何，在立法还是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。^[16]国内外司法裁判对此呈现了分歧：国内微梦诉字节案中^[17]，司法内部就存在认知上的差异：一审法院认为该协议仅是文字层面的宣示，并非技术措施；二审法院虽予以尊重，但也是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《反不正当竞争法》等规定，尤其是新修订的第十三条第三款中关于数据获取行为的具体规定之下，纵观此案Robots协议仅仅只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^[18]，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，其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意义。这种裁判结果的不确定性导致其在实践中陷入无法被真正落实的境地，产生的直接后果是数据源网站无法通过有效法律途径来预防和制止恶意爬取行为，只能诉诸于成本高

昂且容易引发技术封堵手段以维护自身数据安全，这无疑是加剧了网络空间的资源消耗，极大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着系统性缺陷，无法有效回应数据爬取所带来的复杂挑战。一是模糊的法律标准导致可操作性不足。虽然《数据安全法》、《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27条要求开展数据处理活动应当依照法律、法规的规定，但应当如何依照缺乏具体的实施细则与案件指引。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与行政违法性双重认定障碍^[19]，刑法上可能适用非法获取计算机信息系统数据罪，而行政法上过度收集的界限不明，数据合规管理的具体义务与问责机制尚未完全建立，导致企业业务开展时缺乏清晰、稳定的规则指引；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，^[20]合规在很大程度上源于为了规避行政处罚带来的外部压力，而并非提升企业自身核心竞争力、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^[21]。企业未能将数据合规内化为公司治理的有机组成部分，使得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、变化快的行为，传统的侧重于事后处罚与静态合规检查的传统监管显得迟钝且被动。虽然部分企业建立了内控制度，

但是制度的效能不同，整体上缺乏行业共识与执行标准，无法系统地应对数据爬取的全链条风险。

3 数据爬取风险防范的权益均衡进路

前述法律风险与制度困境的根本原因在于数据效率与数据安全、平台自主权与容忍义务之间的权益失衡，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张力与优先序上的抉择。

这种价值上的抉择在学界与司法界引发了讨论，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属性与安全属性，认为未经授权的数据爬取本质上是对个人隐私权、企业财产权和市场竞争秩序的伤害。有学者认为，^[22]爬虫技术加剧了数据控制者与用户之间的权力不平等，应对其适用严格规制。我国司法实践也多采此路径，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与经济价值，主张对公开数据的获取应秉持宽容态度，以避免形成数据垄断，阻碍创新。美国法院在“HiQ诉领英”案^[23]中的判决即

是此路径的极端体现，其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^[24]。

国内外规制理念的差异折射出的是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，需要在效率于与安全动态的均衡中寻找符合实际的举措。数据效的率与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，降低交易成本的同时能够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与应用提供更为丰富的场景和资源，推动安全治理水平的不断提升。对企业而言数据效率与数据安全的均衡固然需承担合规成本，但是对二者的投资能够提供企业的长期竞争力。立法与司法者的智慧在于根据数据类型、应用场景、社会公共利益等因素的不同，因时因地的精细化权衡，在动态发展中寻求特定语境下的最优平衡点。

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问题。《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^[25]。此种权利是激励数据生产

和创新的法律工具，但是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与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^[26]。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，那么容忍义务范围相对广泛；对于后者，平台因投入了实质性劳动与资本，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时需要获取授权，容忍义务范围相对限缩。其次在可爬取的数据范围内，爬取行为的本身也应当受到比例原则的约束^[27]以检验其行为的正当性：目的正当性上，爬取行为需具有合法、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，爬取手段应采取实现该正当目的对平台权益侵害最小的方式，能通过公开 API 接口获取，就不应采用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所追求的利益显失均衡。司法实践中，法院在百度与大众点评案中即运用了类似原理，认定百度的大量抓取行为超过了必要限度^[28]。

数据平台的容忍义务并非无限，其边界取决于数据类型与爬取行为的性质，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通行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，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 建构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套路与执法冲突的前提，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《刑法》第285条的非法获取计算机信息系统数据罪、《民法典》人格权编的侵权责任以及《反不正当竞争法》的规制，法律适用复杂且结果不确定^[29]。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的制度设计，应严格落实《数据安全法》第21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，联合网信、工信、公安等各行业监管机构，吸纳技术专家与法律学者参与，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与识别标准；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与禁止行为，从根本上杜绝因标准不一而导致的管理混乱与法律风险。

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《数据安全法》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^[30]；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，必须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与风险承受能力通过内部评估将其增补进黑名单；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、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，使得数据流通的自由区域也更加明确，从而极大降低企业的合规不确定，清晰的正面指引与负面边界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与应急响应流程，对数据进行监测与预警，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与异常流量监测，设定预警阈值，实现对单IP高频访问、非正常时间访问敏感数据等潜在违规爬取行为的早期发现。一旦发现数据泄露或恶意爬取事件，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，包括黑名单是否未覆盖、权限设置是否过宽等并修订完善，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》等相关法规的要求，正是强调了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 细化数据合规法律标准

破解数据合规尚未落到实地的困境需从立法细化与标准制定两个层面推动原则性法律向具象化、流程化的操作规则转变。首先应当充分发挥现有法律框架下实施指南与国家标准的作用实现监管要求的精细化，《数据安全法》第17条明确规定由国家推进数据开发技术和数据安全标准的体系建设，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

的爬取与跨境流动，处理权限应当仅限于国家授权机构；对于重要数据实行事前评估与审批制，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，获批后方可进行且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，允许合规爬取与利用。其次应重视软法的补充作用，最高人民法院通过发布典型案例、各地监管机关通过发布行政执法指导案例为数据爬取的合规边界提供了具有参考价值的行为指引，这些来自实践一线的司法与执法案例能有效填补成文法的模糊地带^[31]。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，在采集环节，需进行合法性前置审查，确保数据来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27条规定的数据安全管理制度要求；传输与存储环节，必须采取加密、访问控制、去标识化等安全技术措施；使用与销毁环节，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与反向约束两种手段。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相挂钩，对于通过国家认可的合规认证的企业，可以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。强化反向约束方面则严格落实《个人信息保护法》第66条等规定，对违法违规数据爬取行为实施阶梯式惩处，从责令整改、高额罚款，到暂停相关业务、停业整顿直至对严重失信主体实施市场准入限制，罚款必须具有足够的威慑力来确保其金额远高于违规所获利益。同时，建立创新信用评价与披露机制，由市场监管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与赋能者，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，^[32]帮助企业尤其是中小企业理解和落实合规要求，最终形成监管与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，在平台权益与数据流通之间找到平衡点。

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、用户画像等的平台投入实质性资源形成的衍生数据产品，平台享有受法律保护的竞争性权益，其容忍义务范围则大幅收缩，有权要求他人获取授权；第二，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，如尚未设置Robots协议、登录验证、反爬虫技术等，

应被视为其行使自主经营权、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，爬取行为也需目的正当、手段必要、影响均衡，以研究为目的少量爬取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模糊边界得以清晰可操作，而为数据平台的合规管理及第三方的合规爬取提供了稳定的法律预期。

4.3 优化三重授权原则的分类应用

为解决传统三重授权原则在大数据场景下成本过高、阻碍流通的适用困境，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，将数据划分为以下四种类型：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，突破传统的一刀切的严格授权模式，实现数据流转效率的合理化提升。

对于第一类数据，因为涉及到核心个人信息与平台核心资产，必须坚持最严格的三重授权，司法实践在微博诉脉脉案等判例中已对此予以确认。

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收益
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
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,可无需授权	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

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需平台的额外授权,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形成了数据产品,所以爬取仍需获得平台授权,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。^[33]

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现代化发展与精准化适用,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,法理依据在于网站通过发布robots.txt文件已经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了该条款,双方之间这样就成立了一份关于数据抓取规则的合同。这一解释路径完美地契合了《民法典》第471条关于合同订立形式的规定,也为《反不正当竞争法》第2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,由最高人民法院通过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

保障删除权是约束数据爬取行为、贯彻目的限制与存储期限最小化原则的具体落实,企业不能仅被动地响应用户请求,

更需要建立主动的、可验证的数据删除机制。应当将删除权嵌入数据爬取的全流程，企业在规划爬取项目时，就应在设计蓝图（Privacy by Design）中预设数据的存储期限与删除触发条件，项目结束即删除或自爬取之日起满2年自动删除，这不仅是对《个人信息保护法》第19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，爬取获得的个人数据应严格履行《个人信息保护法》第47条规定的删除义务，并采用技术手段确保数据不可恢复；爬取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，以降低合规风险。最后构建可审计的删除验证机制，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证明其已履行删除责任的证据，这将促使删除权从一项抽象的用户权利落地为一项可被监管、可被审计的企业内部管理流程，真正实现彻底的删除。

5 结语

本研究系统研究了企业数据爬取行为的风险、困境与规制路径，创新与贡献主要体现在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；在实践层面所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、司法与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：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的规模、成本收益及企业合规现状进行实证考察使研究结论更具说服力；二是本文立足国内法未充分探讨跨境数据爬取所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是在扼杀技术，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，方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(04):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(05):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,(02):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,(02):165-174.
- [5] 刘莹:《网络爬虫之刑事责任》,载《军法专刊》2022年第4期。
- [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,(02):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(06):18-28.
- [8] 许可.数据爬取的正当性及其边界[J].中国法学,2021,(02):166-188.
- [9] 大数据公司被查背后 网络爬虫侵犯隐私产业链整肃 [EB/OL]. (2019-09-18)[2024-01-20]. https://finance.sina.com.cn/money/bank/bank_hydt/2019-09-18/doc-iicezrq6556039.shtml
- [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/OL]. (2024-03-23) [2024-04-02]. https://www.360kuai.com/pc/920398ecaa668f9fe?cota=3&kuai_so=1&sign=360_57c3bbd1&refer_scene=so_1.
- [11] 观察·人工智能引发的隐私与数据保护风险 [EB/OL]. (2021-09-29) [2024-06-11]. https://www.sohu.com/a/492827067_121123759.
- [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.
- [13] 浙江省杭州市中级人民法院(2018)浙01号民事判决书.
- [14] 广东省深圳市中级人民法院(2017)粤03民初822号民事判决书.
- [15] 上海知识产权法院(2016)沪73民终242号民事判决书.
- [16] Robotstxt.org. 爬虫协议相关内容 [EB/OL]. [访问日期 2024-06-29]. <https://www.robotstxt.org/robotstxt.html>.
- [17] 北京市高级人民法院(2021)京民终281号民事判决书.
- [18] HiQ Labs, Inc. v. LinkedIn Corp., No. 17-16783(2017), paragraph 6.
- [19] 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, (02): 98-108.
- [20] 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.
- [21] 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.
- [22] 同1
- [23] HiQ Labs, Inc. v. LinkedIn Corp., No. 17-16783(2017), paragraph 6.
- [24] 同8
- [25] 王轶. 加快构建数据基础制度, 助推数字经济和数字文明建设 [EB/OL]. 中华人民共和国国家发展和改革委员会官网. [2024-07-22]. https://www.ndrc.gov.cn/xxgk/jd/jd/202212/t20221219_1343657.html.
- [26] 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, (02): 87-102.
- [27] 陈兵, 姚俊羽. 公开数据保护的观念澄清与路径选择[J]. 中国特色社会主义研究, 2024, (02): 38-52.
- [28] 参见上海知识产权法院(2016)沪73民终242号民事判决书.
- [29] 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(06): 37-52.
- [30] 数据安全技术 数据分类分级规则.
- [31] 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(04): 36-45+71.
- [32] 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, (08): 28-34.
- [33] 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, (04): 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，河海大学法学院民商法研究所助理，主要研究方向为经济法、数据法。

收稿日期: XXXX-XX-XX

通信作者:

基金项目: 中央高校基本科研业务费专项资金“新时代我国金融安全法制相关问题研究”(编号: B230207022)

Foundation Items: The Special Fund for Basic Scientific Research Business Expenses of Central Universities: "Research on Issues Related to China's Financial Security Legal System in the New Era" (Project Number: B230207022)