Dispute Settlement 争议解决, 2025, 11(8), 227-234
Published Online August 2025 in Hans. https://www.hanspub.org/journal/ds
https://doi.org/10.12677/ds.2025.118265

Hans汉斯

网络爬取的数据风险及其法律规制研究

曾卓然，梁喜乐，蔡春兰

佛山大学法学院，广东 佛山

收稿日期：2025年7月15日；录用日期：2025年8月13日；发布日期：2025年8月26日

摘　要

数据是竞争与创新的引擎，也是数字技术和数字经济的支撑。作为数据利用的一种重要手段，在缺乏科学合理的规制时，网络爬取在个人隐私、消费者权益以及企业利益等方面潜藏着诸多法律风险。当前，我国网络爬取相关立法呈现出明显的碎片化和滞后性问题，缺乏规制网络爬取行为的具有可操作性的专门性条款。通过借鉴域外相关经验，我国应以分级分类为基础明确网络爬取行为的边界规则，以责任主体为核心构建全链条约束机制，进而完善网络爬取数据风险的法律规制路径。

关键词

网络爬取，数据风险，法律规制

Research on the Risks of Web Scraped Data and Their Legal Regulation

Zhuoran Zeng, Xile Liang, Chunlan Cai

School of Law, Foshan University, Foshan Guangdong

Received: Jul. 15th, 2025; accepted: Aug. 13th, 2025; published: Aug. 26th, 2025

Abstract

Data is the engine of competition and innovation, and also the backbone of digital technology and the digital economy. As an important means of data utilization, web crawling poses potential legal risks to personal privacy, consumer rights, and corporate interests in the absence of scientific and rational regulation. Currently, China's legislation on web crawling is characterized by fragmentation and lag, lacking specialized provisions with operability to regulate web crawling behavior. Drawing on international experiences, China should clarify the boundaries of web crawling behavior based on data classification and grading, and establish a comprehensive constraint mechanism centered on responsible entities, thereby improving the legal regulatory path for web crawling data

文章引用: 曾卓然, 梁喜乐, 蔡春兰. 网络爬取的数据风险及其法律规制研究[J]. 争议解决, 2025, 11(8): 227-234.
DOI: 10.12677/ds.2025.118265

risks.

Keywords

Web Crawling, Data Risks, Legal Regulation

Copyright © 2025 by author(s) and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

1. 引言

在数字经济时代，数据作为核心生产要素的价值日益凸显，网络爬取技术因能高效获取数据而广泛应用，为经济社会发展提供重要支撑。但该技术的失范使用，易引发个人隐私泄露、消费者权益受损、企业利益受侵等风险，扰乱市场秩序。然而，从既有的研究成果来看，中国在网络爬取的规制上仍未构建系统的理论，在司法实践中也未形成较具参考意义的经验。有鉴于此，本文旨在对网络爬取的数据风险进行形成机理解构、规制困境分析和应对路径开辟，以期丰富相关的学理支撑与完善对策。

2. 网络爬取数据风险概述

2.1. 网络爬取的技术解构

网络爬取作为获取网络数据的重要技术手段，其行为主要包括三个阶段即访问、获取和使用阶段[1]。

在访问阶段，公开数据通常对公众开放，允许爬虫访问，不会对用户访问操作进行限制。目前很多网站都能实现大规模爬取，但需注意技术不成熟的小网站，其可能无法承受大规模爬取，易对计算机系统造成破坏。因此，爬取公共数据应避免影响网站正常运行，确保合理访问，也就是爬取数据不可对网站产生任何影响。2023 年颁布的欧盟《数据法案》进一步明确数据访问阶段的合规性要求，强调了数据访问的授权和对数据源系统的保护义务，这为我国相关规制提供了新的借鉴方向[2]。

数据获取是数据使用者使用数据的前提条件。在司法实践中，平台在数据收集和存储上投入了资金和设备，数据获取方需尊重数据平台的利益，即所抓取数据的范围要在授权和公开的边界之内。例如，在微博诉脉脉案中，法院提出“三重授权”规则，要求数据获取方需获得平台和用户授权；在国外脸书诉 Power Ventures 案中，也强调了数据获取方在访问获取数据时，需同时获得用户和脸书平台的授权。这些案例表明，在数据竞争中应尊重平台利益，爬取数据应当限定在目标数据主体授权或公开的范围内活动：抓取用户自行选择公开的数据内容，与普通用户浏览权限等同的部分是正当的；反之，擅自获取受访问控制的数据，则触碰法律红线，可能违反《反不正当竞争法》公平诚信原则，亦可能因为“未经许可擅自获取计算机信息系统数据”而被追究刑事责任。值得注意的是，抓取内容中如涉及公民个人信息，还须遵守《中华人民共和国个人信息保护法》的规定：个人信息处理者可以在合理的范围内处理个人自行公开或者其他已经合法公开的个人信息；个人明确拒绝的除外。个人信息处理者处理已公开的个人信息，对个人权益有重大影响的，应当依照本法规定取得个人同意 [1]。换言之，即便信息公开可见，爬取时也要避免过度收集个人信息，否则可能构成侵害个人信息权益的违法行为。然而，“三重授权”规则在一定程度上限制和阻碍数据流通，如授权流程的复杂性可能导致数据获取的效率降低，且在不同国家和地区的法律框架下，授权的具体要求和执行标准存在差异，这给跨国数据获取带来了挑战。

[1]《中华人民共和国个人信息保护法》第 27 条。

规范爬取数据的使用行为至关重要，需权衡各方利益，包括数据控制方、用户、第三方以及社会利益。具体而言，第三方企业在使用数据时，不能对数据控制方的经营利益造成实质性损害。例如，在大众点评诉百度一案中，百度直接使用其从大众点评抓取的用户信息，导致大众点评用户流失和交易机会减少，削弱其市场竞争力，这种照搬式使用构成不正当竞争。但如果第三方对数据进行深加工，形成具有自身智力成果的产品或服务，这种使用则是被允许的，因为它能为社会带来更大的经济效益。

同时，用户数据涉及隐私，数据使用必须确保用户享有正当的个人信息利益。数据使用者需遵循以下两个方面：一是对个人数据进行匿名化处理，保护用户隐私；二是严格规范数据使用用途，避免侵犯个人隐私或干扰用户正常生活[2]。

2.2. 网络爬取数据的法律风险分析

2.2.1. 对个人隐私的侵犯

数据爬取对个人隐私有着侵犯的风险。其可能未经授权获取个人敏感信息，如联系方式、家庭住址、消费偏好等，这些信息被滥用会干扰个人的正常生活，甚至会带来被不法分子用于诈骗或身份盗窃的风险。

2.2.2. 对消费者权益的冲击

消费者在电商平台、外卖平台和旅游平台等商业活动中的数据极具价值，成为数据爬取者的目标。不良商家或数据中介非法获取这些数据后，可能用于恶意营销，如频繁推送广告或实施“精准宰客”，即利用消费者的个人信息进行价格歧视或其他不公平交易行为，严重损害消费者的合法权益。

2.2.3. 对企业利益的危害

企业投入大量资源收集和分析数据，这些数据是其竞争优势的关键。未经授权的数据爬取会导致企业数据资产流失，竞争对手由此获取到商业秘密、客户名单等敏感信息，从而在市场竞争中获得不正当优势，抢占市场份额，使被抓取企业遭受经济损失。此外，数据抓取还可能破坏企业的商业模式和运营生态，干扰平台的正常数据流动和用户匹配机制，导致用户体验下降、用户流失，进而影响平台的交易规模和盈利水平。

3. 网络爬取数据风险的法律规制困境

3.1. 立法困境

数据作为数字经济发展的新型核心生产要素，其法律保护已成为全球性议题。我国针对网络爬取数据风险的法律规制也做了不少探索，出台了《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等一系列法律文件(具体见表 1)，基于不同的立法目的对数据保护分别作出相应的规定。但从法律规范体系上看，我国现行立法具有分散化、部门化的特点，有关数据的大部分法律条文也仅仅是一般性的原则条款，实际可操作性较低，并不能有效地规制网络爬取数据风险。简言之，我国缺乏规制网络爬取行为的具有可操作性的专门性条款。

Table 1. Legal documents related to regulating the risks of web scraping data
表 1. 规制网络爬取数据风险的相关法律文件

名称	发布时间	性质
《中华人民共和国刑法》	2023 年 12 月 29 日	法律
《中华人民共和国民法典》	2020 年 5 月 28 日	法律

续表

《中华人民共和国网络安全法》	2016 年 11 月 7 日	法律
《中华人民共和国数据安全法》	2021 年 6 月 10 日	法律
《中华人民共和国个人信息保护法》	2021 年 8 月 20 日	法律
《中华人民共和国反不正当竞争法》	2025 年 6 月 27 日修订	法律
《网络数据安全管理条例》	2024 年 9 月 24 日	行政法规
《规范互联网信息服务市场秩序若干规定》	2011 年 12 月 29 日	部门规章

由此可知，目前我国网络爬取相关立法呈现出明显的碎片化和滞后性问题，在面对网络爬取过程中产生的各类数据风险，从而引发个人信息侵犯、数据利用效能降低乃至数据安全威胁等一系列问题时，难以适应数字经济发展需求。

3.2. 司法困境

网络爬取相关立法碎片化、滞后化，暂时无法为网络爬取中的数据保护和利用问题提供全面、有效、合理的保障，此情形下，司法指导路径对于网络爬取数据风险的法律规制就显得格外重要，但也存在一定的局限性，难以有效应对纷繁复杂的网络爬取行为。

目前我国司法实践中，网络爬取案件的裁判标准存在显著模糊性，在不同领域中有不同的表现，其中在民事与竞争法领域尤为明显。

在民事与竞争法领域，司法裁判过度依赖《中华人民共和国反不正当竞争法》中一般条款，法官的自由裁量空间较大，容易导致同案不同判情形的出现，具有较大的不确定性。“酷米客”诉“车来了”案和微博头条数据抓取纠纷案中均强调 roborts 协议的行业惯例效力，依据《中华人民共和国反不正当竞争法》第 2 条该一般性条款认定违反 robots 协议抓取数据具有不正当性[3]。而在“3B”大战中，法院以百度 robot 协议的设置不当为由，认定百度该行为一定程度上限制了数据的有效流通，有碍市场同行合法合规发展，系不正当竞争行为[4]。显然，由于裁判所依据的法律规定较为笼统，法官常常需要在个案中具体问题具体分析，进行个案认定，欠缺稳定性。同时，对个人信息爬取的民事救济不足。在涉及网络爬取的相关案件中，涉案主体通常都是企业，其使用法律的武器来维护企业合法的数据权益[5]。在企业合法的数据权益受到侵害的同时，个人信息、个人隐私也在所难免。但个人信息受损的救济是匮乏的，一方面，个人用户因信息不对称，未能察觉自身的合法权益受到侵害；另一方面，因举证困难，难以证明实际损害等而难以维权。

3.3. 执法困境

行政机关作为网络爬取的主要监管主体，在一定程度上有助于平衡数据的保护和利用间的利益，但目前我国行政监管上依旧存在一些问题，未能进一步提供有效保障。

一方面，网络爬取一般由网信、工信、公安等相关部门进行监管，但《中华人民共和国网络安全法》中并未明确各部门的监管边界，导致“九龙治水”现象，各监管主体常因职责交叉而进行推诿，进而导致行政监管缺位。另一方面，执法技术能力难以匹配网络爬取技术的复杂性。网络爬取主体一般会利用加密、分布式爬虫等技术规避监管，但相关执法部门、执法工作人员往往缺乏相应的技术溯源与证据固定能力，在“车来了”APP 非法爬取公交数据案中，其就需要依赖企业技术协助才能认定爬取行为的破坏性，暴露了行政监管中存在技术短板的问题。

4. 网络爬取数据风险法律规制的域外考究

在数字经济全球化发展的当下，世界各国早已开始有针对性地对数据的保护和利用进行了相关规定。目前，面对我国网络爬取数据风险所引发的一系列问题，我国也正在不断地探索更为完善的法律规制路径，因此特地选取了在数据规制方面比较具有代表性比较成熟的美国和欧盟作为典型进行相关研究。

4.1. 美国的法律规制

美国作为一个数据治理大国，在网络爬取数据风险的法律规制方面已经形成一套具有特色的较为完整的体系，具有一定的借鉴价值。在立法层面，美国现行立法与中国相类似，并没有形成专门性法律，而是通过《计算机欺诈与滥用法》(CFAA)、《反不正当竞争法》及知识产权法进行间接性调整。但美国对网络爬取的法律规制以司法判例为核心的，并已形成了“场景化评估 + 利益平衡”的模式。美国法院在司法裁判中，会注重考察具体场景中的多重客观因素，综合考量和权衡数据主体权益、平台投入回报、社会公共利益等多元利益，才进行判定行为是否正当、合法，以尝试平衡数据保护与数据利用间的利益。目前，美国司法实践中形成了包括动产、计算机欺诈与滥用法和反垄断在内的多元场景化司法规制路径[6]。在个人信息爬取案件中，若爬取方已对个人信息脱敏且未侵犯隐私，法院倾向于认可行为合法性；而对涉及商业秘密的爬取，则严格审查技术手段的正当性。这种“个案分析”模式为数据爬取预留了创新空间，同时也通过“必要设施原则”防止平台数据垄断。但该模式的局限性同样明显，“个案分析”依赖法官自由裁量，易导致裁判标准不统一；且“利益平衡”缺乏量化标准，对平台与爬取方的权益界定易引发争议。

4.2. 欧盟的法律规制

欧盟同样也是数字经济快速发展的大国，形成了与美国“场景化评估 + 利益平衡”的模式较为不同的规制模式，能够为我国网络爬取数据风险的法律规制路径提供一定的思路。欧盟对网络爬取的规制主要以《通用数据保护条例》(General Data Protection Regulation)为核心，严格限制对具有人格性的个人信息的抓取，对基于合法利益或商业目的而进行的数据抓取行为采取较为开放的态度。其一，《通用数据保护条例》第 6 条就明确规定，爬取个人信息必须基于用户的明确同意，且对敏感个人信息的爬取行为进行了额外的限制，在“脸书数据滥用案”中，欧盟委员会认定脸书未经用户同意爬取第三方数据，违反了《通用数据保护条例》，对其进行了巨额罚款。其二，欧盟通过《数据库保护指令》对商业数据集合提供特殊保护，网络爬取主体若未经授权，爬取数据库中的“实质性内容”，则构成侵权。但该特殊保护的客体仅限于“投入实质性投资”的数据库，欧盟对公开数据的爬取并不进行限制，而是采取开放态度。其三，将于 2025 年 9 月 12 日正式开始适用的《数据法案》进一步明确：“非个人数据的合法爬取”可豁免平台授权，但需符合“不损害平台核心功能”的条件。欧盟采取的规制模式一定程度上有利于增强个人信息的保护力度，明确商业数据权益的边界，但也提高了相关企业的合规成本，中小企业易因难以完成用户授权流程而被迫放弃合法爬取，进而限制了数据流通效率。

5. 网络爬取数据风险法律规制的完善路径

5.1. 网络爬取数据的利益冲突

数据爬取带来了多种收益，但其大规模、自动化的数据获取方式也不可避免地触碰到多方权益，形成复杂的利益冲突[7]。本节聚焦数据利用者与数据保护者的核心冲突展开分析。

数据的“流动性”和“复用性”是数字经济的核心驱动力，也即数据的价值在于其可以重复利用、重复开发，基于同一种生产资料可以创造出不同的成果[8]。在社会生活的各项领域内，数据支撑了诸多正当的、有益的实践，如搜索引擎正是依赖大规模网页爬取建立索引，方便公众检索信息[7]；中小企业通过爬取公开数据优化产品策略，实现与巨头的差异化竞争；科研机构爬取公共领域数据推动技术突破；数据聚合平台通过跨领域爬取整合信息，为用户提供更高效的服务。但近年来，随着数据要素市场的深化，利益冲突呈现新形态，商业数据的“公开可访问性”与“竞争性排他性”边界日益模糊，而欧盟《数据法案》(2023 年正式实施)中“数据可携带权”与“反滥用条款”的平衡，也为全球数据爬取争议提供了新的参照框架。

同时，数据的“专属性”和“安全性”诉求愈发凸显，数据的无序流通会导致“投入无回报”“隐私无保障”等不良后果，如平台企业投入大量资源积累的数据，被视为核心资产，担心被爬取后丧失竞争优势；个人用户则担忧爬取行为突破隐私边界(如爬虫绕过授权获取个人通讯记录、消费偏好等，进而用于恶意营销或诈骗)。这种冲突的本质，是数据价值最大化与风险最小化的目标失衡，更深层则是个体权益保护与社会整体福利提升的动态平衡难题——过度限制爬取可能抑制创新，而完全放任则会导致“投入无回报”“隐私裸奔”的双重困境。

因此，网络爬取数据的利益冲突，其本质是数据价值最大化与数据风险最小化的目标失衡，或更进一步地说，是个体权益保护与社会整体福利提升之间的争议。

5.2. 以分级分类为基础，明确网络爬取行为的边界规则

数据的性质(个人/商业/公共)、敏感度、来源不同，其利用与保护的优先级差异显著，决定了其利用与保护的优先级需差异化设定。目前，我国的数据分级分类依据的制度主要有《数据安全法》及相关行政法规、地方性法规和地方政府规章、国家标准、地方标准、行业标准等，我国以《数据安全法》《个人信息保护法》为核心，辅以 2024 年实施的《网络数据安全管理条例》及系列标准构建分级分类框架，但需结合国际最新实践(如欧盟《数据法案》对“工业数据”“公共数据”的特殊规制)动态优化。

5.2.1. 个人数据：坚守网络爬取数据的“最小必要 + 明确授权”底线

对于个人信息，尤其是敏感信息，如生物识别、医疗记录等，无论是否公开，爬取行为必须以“用户明确授权”为前提，且遵循“最小范围、最短存留、特殊用途”的原则。可考虑依据《个人信息保护法》中有关涉及数据行为对个人信息权益影响的相关规定予以治理，譬如，该法第 13 条确立了个人信息处理者合法处理个人信息的基本条件，包括“告知 + 同意”的基本规则，第 44 条赋予了个人对其个人信息的处理享有的知情权、决定权和拒绝权[9]。例如，社交平台用户公开的动态虽可被访问，但第三方若用于画像建模或精准营销，则需单独获得用户授权；非公开的个人数据，应当绝对禁止爬取。

但需注意，“多重授权”(如用户、平台、数据处理者的层级授权)在实践中存在局限性：一方面，用户可能因“授权疲劳”盲目同意，导致权利形式化；另一方面，中小企业因合规成本过高可能规避授权义务。对此，可借鉴欧盟《数据法案》的“授权标准化”经验，探索行业统一授权模板，并建立“授权有效性定期核验”机制，在保障权益的同时降低合规负担。

5.2.2. 商业数据：区分网络爬取数据的“公开性”与“竞争性”，保护投入回报权

对于企业投入资源产生的商业数据，若属于公开可访问但具有竞争性价值的，网络爬取行为需恪守“合理使用”原则，即允许非竞争性利用，禁止竞争性滥用。同时，可以适当引入“数据劳动投入”原则，承认数据收集者对数据投入劳动后的合理利益，同时赋予数据主体(个人)对涉及自身数据的人格权益，并保障社会公共利益对数据的适当利用途径[11]。在此框架内，若爬取方未付出创造性劳动，而是通

过爬取直接获取他人核心商业数据资产，则根据其行为程度落入不同部门法的规制范围内。这也要求了各部门法，如《个人信息保护法》《反不正当竞争法》等，对网络爬取行为和手段作出更为详细的具体规定。

5.2.3. 公共数据：推动数据的应开尽开和有序利用

政府部门、公共机构掌握的公共数据，如气象数据、交通路况、政务服务信息等，属于社会公共资源，应通过统一的开放平台主动提供结构化获取接口，在注明来源的前提下允许无歧视的爬取，但需警惕两类风险：一是敏感公共数据泄露(如应急抢险实时数据被用于恶意炒作)，二是数据被非法牟利(如倒卖政务服务数据)。

对于可能涉及公共安全的较为敏感的公共数据，如应急抢险救灾的实时数据等，可以通过限制爬取主体的资质规制。从某种程度上来说，“官方开放 + 自愿共享”模式丰富了合法的数据获取渠道，在一定程度上削弱了第三方靠数据爬取强取数据的动机[8]。

5.3. 以“责任主体”为核心。构建全链条约束机制

全链条规制需避免“单向从严”或“放任自流”的极端，在权责分配中体现辩证思维。对于网络数据爬取方，应当恪守“合法、正当、必要”原则，这要求爬取方事先评估目标数据用途的合法性，在利用数据时不规避反爬措施、不超过数据使用的合理频率，并在事后保障所爬取数据的安全，采取必要手段(如个人信息去敏化)保证不泄露、滥用数据。

对于网络数据权属方，包括数据生产者、数据管理者、数据处理者等，通过法律的具体规定，明确其明示边界、合理防护的义务。例如，数据聚合平台可通过 robots 协议、用户协议等明确告知数据的开放范围；同时，平台的反爬措施不能以保护为名设置过高的壁垒，避免无理由封禁合规的爬取行为，可引入“反爬措施必要性审查”制度，由监管部门认定措施是否超出“保护必要限度”。

对于网络数据监管方，监管部门应当建立动态执法机制，为爬取主体提供较为合理的容错范围，对窃取商业机密等恶意爬取行为应当严厉打击，而对于中小企业因规则不明确导致的轻微违规等行为，以引导整改为主。同时，可试点“数据爬取沙盒监管”，允许创新型企业在可控环境内测试爬取模式，形成“实践–反馈–规则优化”的良性循环。

6. 结语

网络爬取数据风险法律规制本质是各方利益的平衡，这一平衡并非利用与保护各退一步的静态妥协，而是根据技术发展和社会需求的动态调整：当网络爬取行为过度侵犯隐私或商业利益时，需强化保护规则；当保护措施形成了数据垄断，并进一步抑制创新时，则需放宽合理利用的边界。对网络爬取行为规制的最终目标是让数据成为激活创新发展的工具，而非侵犯权益的利器，让数字经济在安全与效率的共生中持续增长。

参考文献

[1] 张娜. 数据爬取行为的法律边界问题研究[D]: [硕士学位论文]. 北京: 首都经济贸易大学, 2022.

[2] 吴沈括, 柯晓薇. 欧盟《数据法案》的规范要旨与制度启示: 以个人信息保护为视角[J]. 信息通信技术与政策, 2024, 50(1): 2-6.

[3] 李安. 人工智能时代数据竞争行为的法律边界[J]. 科技与法律, 2019(1): 61-70.

[4] 饶传平. 论数据抓取法律风险的流程化管理[J]. 东方法学, 2023(6): 28-42.

[5] 张依. 数据抓取行为的法律规制研究[D]: [硕士学位论文]. 西安: 西北大学, 2022.

[6] 陈顺. 网络爬虫抓取数据法律规制的不足与完善[D]: [硕士学位论文]. 南京: 南京大学, 2021.
[7] 李谦. 美国场景化司法规制数据爬取的经验与启示[J]. 电子政务, 2020(11): 86-98.
[8] 常烨. 生成式人工智能数据“投喂”的著作权侵权行为规制[J]. 科技与法律(中英文), 2025(2): 31-41.
[9] 任浏玉. 公开商业数据爬取行为的规制路径[J]. 知识产权, 2022(7): 110-126.
[10] 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.
[11] 付新华. 论数据治理的使用权范式[J]. 中外法学, 2024, 36(6): 1584-1600.