

全球数据治理竞合下爬取行为规制的中国路径

杨海楠, 刘海芳

合肥工业大学文法学院, 安徽 合肥

收稿日期: 2025年3月28日; 录用日期: 2025年4月21日; 发布日期: 2025年4月30日

摘要

随着数字经济的蓬勃发展,数据爬取技术及其相关产业链迅速兴起,成为两个复杂领域交汇的产物。1990年搜索引擎的出现标志着数据爬虫技术的首次应用,随后在1993年首个数据爬虫诞生,紧接着1994年Google、Yahoo等成熟搜索引擎相继问世,发展速度令人瞩目。在大数据时代,数据被视为重要的生产力,数据爬虫与反爬虫之间的对抗也日益激烈。然而,数据爬取技术的迅猛发展使得管理和规制的成本与难度显著增加。尽管数据爬取技术本身是中立的,但其使用行为可能存在不当或违法的风险,进而带来诸多负面后果。因此,亟需通过完善相关法律法规、加强技术监管以及优化数据治理机制,以平衡技术创新与数据安全,确保数字经济时代的健康与稳定发展。

关键词

数据爬取, 个人信息保护, Robots协议

Legal and Ethical Considerations and Countermeasures for the Use of Intelligent Care Devices by Disabled People Living Alone

Hainan Yang, Haifang Liu

School of Humanities and Law, Hefei University of Technology, Hefei Anhui

Received: Mar. 28th, 2025; accepted: Apr. 21st, 2025; published: Apr. 30th, 2025

Abstract

With the development of technology, intelligent care devices have brought great convenience to the lives of disabled people living alone. However, they have also given rise to a series of legal and ethical

issues. This paper delves into the legal and ethical issues involved in the process of disabled people living alone using intelligent care devices, including privacy protection, data security, right to autonomous decision-making, and liability definition. It also puts forward targeted countermeasures such as improving laws and regulations, strengthening technological safeguards, and enhancing ethical education. The aim is to promote the legal, compliant, and ethical application of intelligent care devices while safeguarding the rights and interests of disabled people living alone.

Keywords

Web Crawling, Personal Information Protection, Robots Exclusion Protocol

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. 数据爬取行为的概述

1.1. 数据爬取的性质及应用

数据爬取技术是中立的, 数据爬取行为的正当性取决于此项技术的使用者爬取行为是否无害、爬取数据是否侵权等。一部分学者认为数据爬取行为属于不正当竞争行为[1], 而一部分学者认为应当否定将数据爬取行为定义为不正当竞争行为。数据爬取技术最初被应用于搜索引擎页面, 后出现数据爬取者将所爬取的数据搬运到其网页当中, 运用深度链接技术使用户获取目标数据, 即用户点击网页会跳转到第三方网站(被爬取者)。近来运用广泛的生成式人工智能也少不了数据爬取的运用, 其需要不断的对公共数据爬取、分析及训练[2]。此外, 数据爬取也有被滥用的局面, 例如在金融行业, 虽然银行采取大数据风控措施, 但其水平尚浅仅仅只是将数据线上化, 并未真正有效实现管控。例如数据爬取技术被用于获取用户个人信息的, 将这些个人信息打乱、重新排列组合后出卖以此谋取利益, 构成破坏计算机信息系统罪。

1.2. 数据爬取行为的分类与特征

数据爬取行为的复杂性要求对其进行多维分类, 以便精准制定规制措施: 第一, 按数据来源分类, 分为公开数据和受限数据[3]。公开数据指社交媒体、新闻网站等可公开访问的数据, 需遵守网站协议, 如 robots 协议。受限数据有用户个人信息、企业商业秘密等, 需获得明确授权。第二, 按爬取目的分类。商业用途目的(市场分析、竞争情报), 需防范不正当竞争; 学术研究目的, 需确保数据使用的非营利性和透明度; 个人使用目的, 需限定数据规模与用途。第三, 按技术手段分类。分为 API 接口式爬取, 需遵守接口调用频率限制; 网页解析式爬取, 需避免破坏网页结构或侵犯知识产权; 自动化脚本工具, 需防止绕过反爬虫技术[4]。

1.3. 数据爬取合法性判定

人工获取数据与爬虫获取数据的区别在于数据爬虫获取的数据量更加庞大、数据爬虫更快捷, 但此并非是给其定罪的依据, 数据爬虫入罪的一重大原因是其在批量数据爬取过程中未能向个人信息权利人、代管人发出请求同意的申请, 或者是利用爬虫程序过程中造成计算机系统损害的结果。数据爬虫能够发展至今是时代的需求, 如今对其进行规制也是时代的需求, 究竟如何进行规制, 众说纷纭。过往对数据爬取的入罪要求极低, 只需存在该恶意行为, 近来学者们提出应提高数据爬取入罪要求, 不仅对恶

意数据爬取行为定性, 而且对恶意数据爬取行为定量, 更精准的对数据爬取进行打击。

第一, 合法性原则。传统的数据爬取侵权的判定聚焦于数据爬取行为, 也即判断数据爬取行为是否具有技术侵入性[5]。除根据国家和地区关于数据保护、知识产权、网络安全法等相关法律法规来判定外, 数据爬取侵权行为的判定还需尊重网站使用条款。第二, 授权原则。数据爬取应当有着明确授权或者合理推定性授权。明确授权这种授权可以是书面的、电子的或者通过网站提供的应用程序编程接口(API)来获取数据。第三, 正当目的原则。该原则指爬取来的数据要基于非商业目的且合理使用, 如果是出于个人学习、研究、新闻报道等非商业目的的数据爬取, 并且没有对数据所有者造成不合理的损害, 在一定程度上可能被视为正当的[6]。第四, 损害原则。数据爬取侵害行为最有可能侵害的是数据所有者利益, 需对数据所有者权益的损害评估。如果数据爬取行为导致数据所有者的经济利益受损, 同时也需关注到爬取者的合法权益, 也就是需要平衡数据使用者的权益。在判定侵权时, 考虑数据爬取者的权益, 如果数据爬取者是基于合理的目的, 如开发新的服务或产品, 并且没有过度损害数据所有者权益的情况下, 可能需要在两者权益之间进行平衡。

2. 数据爬取规制的现状与困境

如前所述, 数据爬取侵权可能涉及侵犯公民个人信息罪、侵犯著作权罪, 这些罪名的存在从侧面反映出数据爬取领域的乱象与风险。然而, 法律并非仅仅止步于罪名的设立, 规制体系也在不断发展完善以应对数据爬取活动中的各类情况[7]。当下, 数据爬取规制在实践中呈现出复杂的现状与面临诸多困境, 亟待深入探究。本章将探讨在部门法框架下对数据爬取规制的现状以及完善规制体系所面临解决的问题。

2.1. 数据爬取的规制现状

在民法方面, 对于数据爬取的规制主要涉及数据隐私和数据侵权的规定。有关数据隐私权益的民法保护, 在《民法典》第 127 条, 其是一个宣示条款, 将数据放在民事权利一章, 表明数据权益本身是一项民事权益, 这不仅为数据纳入民法保护范围提供了依据, 也为单独立法预留了空间。数据权益涉及多方面, 包括个人信息、隐私等, 也与交易、共享相关[8]。司法中也可根据《中华人民共和国民法典》第一千一百六十五条的规定进行数据爬取是否侵权的判定, 行为人因过错侵害他人民事权益造成损害的, 应当承担侵权责任。数据爬取侵权可能会触犯刑法, 例如构成侵犯公民个人信息罪、侵犯商业秘密罪和破坏计算机信息系统罪等, 司法解释方面, 目前并没有专门针对数据爬取的特定司法解释。但在司法实践中, 通常会依据《刑法》相关条文来对涉及数据爬取的行为进行认定和处理, 比如《刑法》第二百八十五条规定的非法获取计算机信息系统数据罪等[9]。通过刑法规制, 可以对这些违法行为进行严厉打击, 维护社会公平正义和正常的经济秩序。具体而言可依照侵犯法益的不同, 对具体的数据爬取行为进行不同规制。例如, 大众点评诉百度案中百度通过爬虫技术抓取大众点评用户评论数据用于自身产品, 法院认定其构成不正当竞争[10]。而在其它案件中, 第三方通过恶意爬虫非法获取用户个人信息并贩卖, 最终以“侵犯公民个人信息罪”定罪。行政监管部门对网络爬虫程序的合法性、合理性以及安全性等方面进行审查和管理, 确保程序不侵犯他人合法权益。面对数据爬取侵权, 行政监督管理部门会采取多种措施进行打击, 包括第一责令停止侵权行为, 一旦认定存在侵权, 会立即要求侵权方停止数据爬取行为, 避免侵权损害的进一步扩大。

2.2. 数据爬取规制面临的困境

虽数据爬取侵权行为在部门法领域下能得到较完善的规制, 但当前的数据爬取规制体系仍会面临一些问题, 比如数据爬取规制在各部门法之间的衔接以及具体条款适用上可能存在一些争议。因为不同部门法对数据爬取行为的规制角度和力度有所不同, 所以导致部门法的衔接上出现不协调的问题[11]。司法

实践中规制数据爬取会存在一些难点, 会面临证据收集困难的问题, 因为数据爬取行为往往较为隐蔽, 难以直接获取到爬取的过程和相关证据, 所以确定爬取行为的主体、时间、方式以及所获取数据的范围和用途等都存在挑战, 这使得在法庭上证明数据爬取行为的存在及违法性变得困难[12]。即使收集到数据爬取违法的证据, 在数据是否违法的判断、行为的性质和损害的认定等问题上也有困扰[13]。数据爬取行为可能涉及多种法律关系和法律问题, 如可能同时涉及反不正当竞争、侵犯知识产权、侵犯商业秘密、违反计算机信息系统安全保护规定等多个领域, 行为性质的复杂性增加了司法认定的难度。

3. 全球数据治理下爬取行为规制的域外经验借鉴

3.1. 欧盟：严格保护导向的规制体系

欧盟在数据治理方面有较为丰富的经验和立法实践, 在数据爬取规制方面建立了一套明确且完善的制度, 该制度从多个维度发挥着重要作用, 对数据的合理利用、市场活力的激发以及竞争秩序的维护都有着深远意义。欧盟数据爬取规制制度清晰地明确了数据访问、使用等相关规则, 其核心目标是确保用户能够对自身的数据拥有更多的控制权[14]。同时, 欧盟数据爬取规制还引入了数据访问权利, 保障每一个用户, 无论是个人还是组织, 都拥有获取和产生数据的权利, 而且这种权利涵盖了个人数据和非个人数据[15]。这一规定打破了数据流通的壁垒, 促进了数据在不同主体之间的合理流通和共享, 为数据的广泛应用奠定了基础。且欧盟数据爬取规制制度通过建立数据共享协调框架, 积极推动跨部门的数据共享和开发利用。在这个框架下, 不同部门之间的数据能够得到更有效的整合和分析, 有助于深入挖掘数据背后的潜在价值。

3.2. 美国：市场自律与场景化监管模式

在当今数字化时代, 数据爬取活动日益频繁, 其规制制度的完善至关重要。美国在数据爬取规制制度方面积累了一些值得参考的经验, 这些经验从不同维度为数据爬取的规范管理提供了思路。美国主要采取市场自律与场景化监管模式, 场景化司法规制着重考察具体场景中的多重客观因素[16]。比如, 在某些特定的商业场景中, 数据爬取可能是为了进行市场调研, 以帮助企业了解竞争对手的动态和市场趋势。而在其他场景下, 数据爬取可能是用于学术研究, 以推动科学知识的进步。目前, 美国已经形成了包括动产、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。在动产方面, 当数据被视为一种具有经济价值的动产时, 对其爬取行为可能会受到动产相关法律的规制[17]。计算机欺诈与滥用法则主要针对那些通过非法技术手段进行数据爬取的行为。而反垄断方面的规制则关注数据爬取行为是否会对市场竞争造成不良影响。如果一家企业通过大规模的数据爬取来垄断市场信息, 从而限制了其他竞争对手的发展, 那么这种行为可能会受到反垄断法的制裁。

3.3. 其它国家有代表性的规制模式及经验借鉴

除了欧盟凭借《通用数据保护条例》(GDPR)在数据爬取规制领域树立起具有广泛影响力的标杆, 以及美国通过一系列联邦和州层面的法律、法规对数据爬取行为进行规范之外, 澳大利亚、加拿大和英国等国家同样在数据爬取规制方面积累了宝贵且值得借鉴的相关经验。澳大利亚有专门的数据保护法规, 对个人信息的收集、使用和披露等进行严格规范, 包括对数据爬取行为的约束, 以保护公民的隐私和数据安全。澳大利亚能够结合数字经济发展的实际情况, 对数据爬取过程中的数据收集、使用和共享等环节制定了细致且具有可操作性的规则, 具有完善的隐私法律框架。

加拿大重视数据爬取的规制, 其通过专门的法律法规和监管机构, 对数据爬取行为进行严格监督和管理[17]。不仅对可公开访问个人数据的保护有明确规定和举措, 而且在平衡数据创新利用和个人隐私保

护之间取得了良好的成效。英国的监管机构信息专员办公室会对数据爬取等行为进行监管, 依据数据保护和隐私法, 要求社交媒体和网站运营商保护个人信息不被非法抓取, 在数据爬取的合法性界定和隐私保护方面有相关实践和规定。这些国家的经验在数据保护原则、法律框架、监管机制等方面都可能对我国有一定的借鉴意义。

4. 基于域外经验的数据爬取法律规制体系完善

4.1. 平衡借鉴：欧盟权利保护范式与数据流通需求的兼容

4.1.1. GDPR 在中国的具体适用路径

其一, 构建合规评估机制。要求企业提交《数据爬取影响评估报告》, 涵盖数据来源、处理方式及风险防控措施; 借鉴欧盟“充分性认定”机制, 我国建立数据爬取行为的合规评估体系十分必要[18]。该体系一方面要保障数据主体的权利, 让数据爬取行为透明化, 使数据主体清楚自己的数据被如何使用; 另一方面, 也要促进数据的合法流通。数据作为一种重要的生产要素, 合理的流通能够推动经济的发展和 innovation。建立合规评估体系需要多方面的努力。其二, 监管部门定期抽查。对未合规企业实施“阶梯式处罚”[19]。政府应出台相关政策和法规, 明确数据爬取的规则 and 标准; 企业要加强自律, 遵守法律法规, 积极开展合规评估; 同时, 公众也应提高数据保护意识, 维护自身合法权益。其三, 建立数据分级分类制度。将数据分为公开级、受限级、机密级, 分别设定爬取权限, 如受限数据需双重授权。通过以上三项使得数据爬取在法律框架内进行, 这有效避免了数据的滥用和对个人隐私的侵犯。通过多方协作, 构建一个健康、有序的数据爬取环境, 既能充分发挥数据的价值, 又能切实保护数据主体的权利, 为我国数字经济的可持续发展奠定坚实基础。

4.1.2. 权利保护与数据流通的平衡

在数据爬取规制方面, 首要任务是明确数据权利的范围和边界。这是因为若不加以明确, 可能会出现过度限制数据流通的情况, 阻碍信息的有效传播与利用。可以借鉴欧盟提出的“数据可携权”理念[20]。在符合规定的情况下授权爬取其数据时, 能使得数据资源得到更充分的利用。另外, 建立数据分级分类制度也是关键举措。不同类型的数据具有不同的性质和敏感度, 公开数据与受限数据在爬取规则上应有所区别。对于公开数据, 可适当放宽爬取规则, 以提高数据流通的效率, 让更多有价值的信息能够快速传播和共享。而对于受限数据, 则要制定严格的爬取规则, 确保数据的安全性和隐私性不被侵犯。通过这种差异化的爬取规则, 可以在保障数据流通效率的同时, 最大程度地保障数据安全。

4.2. 技术适配：美国场景化监管的本土化改造路径

4.2.1. 美国场景化监管的实践

美国通过《计算机欺诈与滥用法》(CFAA)等法律, 构建了数据爬取行为的场景化监管体系。该体系根据不同场景设定不同的法律义务, 以平衡数据爬取的合理需求与数据所有者的权益保护。在商业竞争场景中, 数据爬取可能成为企业获取竞争优势的手段。一些企业为了了解竞争对手的产品信息、市场策略等, 会通过数据爬取收集相关数据。然而, 这种行为必须在法律框架内进行。法律要求企业不能使用非法手段绕过对方的数据保护机制, 不能将爬取的数据用于不正当竞争行为, 如恶意诋毁竞争对手、窃取商业机密等。一旦违反这些规定, 企业将面临法律的制裁, 包括巨额罚款和刑事处罚。在学术研究场景下, 数据爬取则是为了推动知识的进步和学术的发展。研究人员为了开展实验、分析数据等, 可以在一定条件下进行数据爬取。但他们需要遵守严格的规定, 如明确告知数据来源、仅将数据用于学术研究目的、不能进行商业盈利活动等。这种场景化的监管, 既保障了学术研究的顺利进行, 又防止了数据的

滥用。美国的“合理使用”原则为数据爬取行为提供了一定的灵活性。该原则允许在特定情况下,对数据进行合理的使用,而不构成侵权。我国可以借鉴这一原则,在数据爬取规制中引入场景化分析,明确合法爬取行为的边界。通过对不同场景下的数据爬取行为进行细致分析,制定相应的规则,既能保护数据所有者的权益,又能促进数据的合理利用。

4.2.2. 本土化改造路径

不可一味移植国外的经验,需要结合我国的实际情况,对数据爬取监管进行本土化改造。首先,制定适用于本土的数据爬取技术标准至关重要。我国的数据爬取技术主要包括 API 接口式爬取和网页解析式爬取。对于 API 接口式爬取,应设定严格的接口使用规范。爬取主体需要遵守接口的调用频率限制,确保不会对数据所有者的系统造成过大压力。同时,要保证爬取的数据仅用于合法目的,不能进行非法的数据交易或其他不当行为。对于网页解析式爬取,要防止爬取主体通过恶意脚本破坏网页的正常运行,避免侵犯网页的知识产权。此外,爬取主体应尊重网站的“机器人协议”,遵守网站设定的访问规则。

其次,建立数据爬取行为的备案制度是保障监管灵活性与有效性的重要举措。在我国,大规模的数据爬取行为可能会对数据所有者和社会产生较大影响。因此,要求爬取主体在特定场景下,如大规模爬取时,向监管部门备案是必要的。备案内容应包括爬取的目的、范围、数据来源、使用方式等详细信息。监管部门可以根据备案信息对爬取行为进行实时监控和评估,及时发现并处理违规行为。同时,备案制度也有助于监管部门了解数据爬取行为的整体情况,为制定科学合理的监管政策提供依据。

4.3. 系统建构:亚太地区跨部门协作机制的启示

亚太地区在数据爬取监管的跨部门协作上已经取得了一定的成果。以日本为例,其通过《个人信息保护法》与《数据利用促进法》的协同,构建了数据爬取行为的双重监管框架。这种方式兼顾了数据权利保护与数据利用促进,既确保了个人信息不被非法侵犯,又为数据的合理开发利用创造了良好环境。新加坡则通过《个人数据保护法》与《网络安全法》的联动,建立了数据爬取行为的跨部门协作机制。这一机制使得不同部门在监管过程中能够相互配合,确保了监管的一致性与高效性,避免了因部门之间协调不畅而导致的监管漏洞。借鉴亚太地区的协作经验,我国在数据爬取监管领域的跨部门协作机制构建也势在必行。

首先,建立数据爬取行为的联合监管机制至关重要。当前,我国涉及数据爬取监管的部门众多,包括网信办、市场监管部门、公安机关等。然而,各部门之间职责分工不够明确,容易出现监管重叠与空白的情况。通过建立联合监管机制,能够清晰划分各部门的职责,使各部门在监管过程中各司其职、协同合作^[21]。例如,网信办可以负责对网络数据的整体统筹和宏观监管,市场监管部门则专注于对数据爬取行为的市场规范和商业伦理监管,公安机关则着重打击数据爬取中的违法犯罪行为。这样一来,整个监管体系将更加严密,监管效率也将得到显著提升。

其次,推动数据爬取行为的标准化建设是提升监管科学性与透明度的关键。目前,我国在数据爬取方面缺乏统一的技术规范与执法标准,导致不同地区、不同部门在监管过程中存在尺度不一的问题。制定统一的技术规范,能够明确数据爬取的合法边界和技术要求,使企业和从业者清楚知道哪些行为是被允许的,哪些是被禁止的。同时,统一的执法标准能够确保监管的公正性和一致性,避免因标准不一而引发的争议和不公平现象。亚太地区在数据爬取监管的跨部门协作方面为我国提供了有益的启示。我国应积极借鉴这些经验,加快构建适合我国国情的数据爬取监管跨部门协作机制,通过建立联合监管机制和推动标准化建设,提升我国数据爬取监管的水平,保障数据安全和个人权益,促进数字经济的健康发展。

参考文献

- [1] 何宗文. 论数据爬取的反不正当竞争法规制[J]. 兰州文理学院学报(社会科学版), 2025, 41(1): 88-93.
- [2] 胡佳烁. 互联网公开数据爬取行为的反不正当竞争规制研究[J]. 网络安全与数据治理, 2024, 43(12): 74-79.
- [3] 陈淑婷. 中小企业爬取数据的正当性及规制路径[J]. 华南理工大学学报(社会科学版), 2024, 26(4): 100-110.
- [4] 陈墨超. 恶意爬取企业数据行为的刑法规制研究[D]: [硕士学位论文]. 南京: 南京信息工程大学, 2024.
- [5] 洪翔. 网络数据爬取行为的入罪边界研究[D]: [硕士学位论文]. 兰州: 西北师范大学, 2024.
- [6] 丁稳. 数据爬取不正当竞争认定的反思及路径重构[J]. 东华理工大学学报(社会科学版), 2024, 43(2): 165-174.
- [7] 曹岚欣. 恶意爬取数据行为的刑法规制边界——以非法获取计算机信息系统数据罪为视角[J]. 中国石油大学学报(社会科学版), 2024, 40(2): 130-138.
- [8] 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.
- [9] 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.
- [10] 贾振宇. 论反不正当竞争法视角下数据爬取行为的裁判路径与完善[J]. 胜利油田党校学报, 2023, 36(6): 41-46.
- [11] 李雪健. 平台数据爬取行为的刑法规制[J]. 刑法论丛, 2022, 70(2): 235-260.
- [12] 李泽, 雷菲萍. 数据权属的界定——以企业间数据爬取为例[J]. 知与行, 2023(5): 76-84.
- [13] 王小源, 姚万勤. 企业间数据爬取行为的竞争法规制逻辑与完善路径[J]. 长江论坛, 2023(5): 68-78.
- [14] 李卓君. 爬取公开数据行为的刑法规制[D]: [硕士学位论文]. 重庆: 西南政法大学, 2023.
- [15] 伊温和. 数据爬取行为的反不正当竞争法规制研究[D]: [硕士学位论文]. 上海: 华东政法大学, 2023.
- [16] 孙悦. 论爬取公开数据行为的刑法规制与入罪限度[D]: [硕士学位论文]. 北京: 北京交通大学, 2023.
- [17] 曹欣欣. 数据安全视角下数据爬取行为的刑法规制研究[D]: [硕士学位论文]. 武汉: 中南财经政法大学, 2023.
- [18] 任炳恒. 企业间数据爬取纠纷的司法裁判研究[D]: [硕士学位论文]. 武汉: 中南财经政法大学, 2023.
- [19] 危冰怡. 我国数据爬取行为的反不正当竞争法规制研究[D]: [硕士学位论文]. 西宁: 青海师范大学, 2023.
- [20] 王晓妍. 网络数据爬取行为的刑法规制研究[D]: [硕士学位论文]. 石家庄: 河北师范大学, 2023.
- [21] 李思阳. 中美数据爬取司法案例比较研究[J]. 法学评论, 2024(3): 45-58.