网络爬虫行为的罪责认定路径:数据确权与利益平衡

王华伟

内容提要:确定网络爬虫的刑事责任边界,核心问题在于对数据获取行为是否取得授权的判断。关于数据犯罪中的"未经授权"或"超越授权",美国判例先后存在代理范式、合同范式、撤销范式、代码范式等不同认定路径。上述范式背后所代表的合约权利标准和技术障碍标准各有利弊,二者并非互斥择一,而应是递进互补的关系。网络爬虫刑事责任的认定,应当首先通过场景式、类型化的思路进行数据确权分析。在此前提下,合约权利的违反奠定了数据爬取行为的基础不法。对合约的形式应当进行限定,其中爬虫协议发挥着举足轻重的作用。在此基础上,技术障碍的突破进一步提升和确证了刑事不法,由此可以限制处罚范围的过度扩张。技术障碍的认定不宜过度严苛,典型的反爬措施可以归入此类,对此应当妥善考虑企业数据权利、平台运营模式等多重利益的平衡。

关键词:网络爬虫　数据确权　爬虫协议　反爬措施

王华伟,北京大学法学院助理教授。

伴随着新一轮科技革命的推进,数据成为当代社会基本的生产和生活要素。数据的高效、合理、有序的流通已经成为社会向前发展的重要推动力量。然而,面对网络空间海量的数据,包括个人和企业在内的社会主体能够在何种程度上自由获取,需要在法律上予以明确。尤其是借助网络爬虫这一技术手段来批量抓取数据是否应当承担法律责任的问题引发了理论与实践的极大争议。网络爬虫是一个自动下载网页的计算机程序或自动化脚本,通常从一个URL(即网络地址)集合开始运行,将这些URL放入一个有序的待爬队列里,按照一定顺序从中取出URL并下载所指向页码的内容。〔1〕网络爬虫本身所具有的积极社会效用显而易见。一方面,网络爬虫是网络空间抓取数据的常用工具,在许多领域被广泛运用;另一方面,对数据权利人而言,网络抓取也越来越多地扮演一种互惠性的角色,因为其能够在一定程度上促进数据本身利用的优化。〔2〕

〔1〕 参见孙立伟、何国辉、吴礼发:《网络爬虫技术的研究》,《电脑知识与技术》2010年第15期,第4112页。

〔2〕 See Jeffrey Kenneth Hirschey, Symbiotic Relationships: Pragmatic Acceptance of Data Scraping, 29 *Berkeley Technology Law Journal* 897, 919-922 (2014).

然而,网络爬虫的使用者和数据支配者(数据权利人)经常产生冲突,继而引发诉讼。使用网络爬虫涉及民事、经济争议(尤其是不正当竞争)的判例早已有之,竞争法、行政法等领域已经展开了一定研究。值得注意的是,近年来在我国司法实践中逐渐出现了一系列通过网络爬虫获取数据入刑的案例。鉴于网络爬虫所具有的积极社会价值,将利用该技术抓取数据的行为宽泛地予以刑事处罚显然并不妥当。这不仅没有正当性和必要性,而且会严重挤压数据自由流通的空间。但是,如何确定网络爬虫行为罪责认定路径,亟待予以系统理论回应。在诸多争议中最为重要的问题在于,采取何种理论范式,对爬取数据行为是否授权(或是否超越授权)进行分析。对此,我国刑法学界尚未进行深入探讨。本文将侧重对美国比较法的研究,通过梳理和分析相关理论范式和认定标准,结合场景化、类型化的数据确权思路,综合考虑多方主体利益的平衡,探讨合理的实质授权标准,为网络爬虫行为的罪责认定提供妥当的参考思路。

一 规制结构与理论范式

网络爬虫行为的数据授权问题在不少国家都引发了讨论。一方面,授权问题的讨论取决于立法层面对数据犯罪所采取的规制模式;另一方面,其又受制于司法层面所依赖的理论范式。只有结合这两个层面的考察,才能深层透视爬取网络数据行为的核心争议。

(一)立法模式与规制结构

在一些代表性国家和地区的立法框架中,大都设置了专门的数据犯罪条款来规制对数据的非法访问和获取,据此可以相应处理未经授权的数据爬取行为。对于作为数据犯罪构成要件的手段条件的界定将直接影响爬取数据行为的刑事责任边界。美国《计算机欺诈与滥用法》(*The Computer Fraud and Abuse Act*, CFAA)对非法访问、获取数据行为进行了细致的规定。按照该法,未经授权或超越授权故意访问计算机并通过这种手段获取特定数据的行为构成犯罪。这一立法一方面将实质成立条件从“未经授权”扩展至“超越授权”;另一方面并未设置明确的技术手段(如突破安全措施)限定构成要件的适用范围,显然语义辐射面相当宽阔,这也为司法适用上的极大争议埋下伏笔。与此类似,我国《刑法》第 285 条第 2 款规定,违反国家规定,侵入计算机信息系统或者采取其他技术手段,获取计算机信息系统中数据的,构成非法获取计算机信息系统罪,这显然也是一种颇为模糊且宽泛的立法。此外,2011 年最高人民法院、最高人民检察院发布的《关于办理危害计算机信息系统安全刑事案件应用法律若干问题的解释》第 2 条对“专门用于侵入、非法控制计算机信息系统的程序、工具”进行了定义,其中明确采用了“未经授权或者超越授权获取计算机信息系统数据”这样的表述。由此可见,在数据犯罪的立法模式和规制结构上,美国和中国的做法都形成了较大的不确定性,这就需要在法律解释层面通过适当的分析路径和理论范式予以限定。

(二)判例发展与理论范式

美国早在 1986 年便通过了《计算机欺诈与滥用法》,而我国直到 2009 年《刑法修正案(七)》才增设了非法获取计算机信息系统数据罪,美国理论界与实务界对非法访问、获

取数据行为的探讨较之于我国更为深入,值得参考。通过对各种不同立场典型判例的梳理,理论上总结出了一系列法律适用标准和理论范式。例如,有的学者将无授权访问的界定归结为五种范式,[3]有的学者提出了相对简化的三重标准,[4]还有的学者在三重标准之外,进一步提炼出第四个撤销范式,[5]以上不同的标准和范式实际上存在一定程度的相互交叉和重叠。本文大体结合判例立场的时间发展顺序,选取几种代表性的方案加以介绍与评析。

1. 代理范式

所谓代理范式,通常是指企业员工违背职业忠诚,利用其职业上的便利访问和获取数据,因而被认为属于"未经授权"。在 Shurgard Storage Centers, Inc. v. Safeguard Self Storage, Inc. 案(下称"舒加德案")中,原告舒加德公司的员工利兰通过电子邮件,将其可以访问的原告公司的商业机密和专有信息发送给了被告公司。法院认为,当该员工已经成为被告公司代理时,他在原告公司的职权就已终止,所以成立"未经授权"。[6] 在类似的 International Airport Centers, LLC v. Citrin 案中,被告西特林是受雇于原告公司的一名员工,负责对目标收购资产进行调查,后西特林决定"自立门户",于是在公司曾借给他办公的电脑中擅自安装了删除数据的软件。法院援引了舒加德案判决,认为西特林违背了员工的忠诚义务,因而也构成"未经授权"。[7]

在理论上,相当有力的观点认为,代理范式将行为人的动机作为判断依据,对是否授权的理解极为宽泛,大大拓展了处罚范围,[8]但是这一标准无法准确回答哪些员工行为是被禁止的,因此有违宪法上"不明确即无效原则"。[9] 虽然代理范式所涉及的上述案件与爬取网络数据案件存在明显差异,数据爬取者和数据控制者之间通常并无雇佣或代理关系,但是如果采取代理范式所侧重的主观标准,由于网络爬虫的使用往往违背数据控制者的授权意愿,那么照此逻辑推导,数据爬取者入罪的可能性相当之高。

2. 合同范式

所谓合同范式,并非完全统一的标准,一般是将对一定合同条款、使用条款、访问规定的违反作为授权与否的主要认定依据。在 United States v. Drew 案中,被告人德鲁违背社交网站 MySpace 的使用条款,使用他人照片假冒注册为 16 岁男孩,对一名 13 岁少女进行欺骗和骚扰,导致后者自杀。尽管陪审团认为被告有罪,但是法官认为单纯违反网站使用

---

〔3〕 See Patricia L. Bellia, A Code-Based Approach to Unauthorized Access under the Computer Fraud and Abuse Act, 84 *George Washington Law Review* 1442, 1446-1457 (2016).

〔4〕 See Andrew Sellars, Twenty Years of Web Scraping and the Computer Fraud and Abuse Act, 24 *Boston University Journal of Science and Technology Law* 372, 393 (2018).

〔5〕 See Annie Lee, Algorithmic Auditing and Competition under the CFAA: The Revocation Paradigm of Interpreting Access and Authorization, 33 *Berkeley Technology Law Journal* 1307, 1312-1326 (2018).

〔6〕 See Shurgard Storage Centers, Inc. v. Safeguard Self Storage, Inc. 119 F. Supp. 2d 1121, 1125 (W. D. Wash. 2000).

〔7〕 See International Airport Centers, LLC v. Citrin, 440 F. 3d 418, 421 (7th Cir. 2006).

〔8〕 See Orin S. Kerr, Cybercrime's Scope: Interpreting Access and Authorization in Computer Misuse Statutes, 78 *New York University Law Review* 1596, 1633-1634 (2003).

〔9〕 See Orin S. Kerr, Vagueness Challenges to the Computer Fraud and Abuse Act, 94 *Minnesota Law Review* 1561, 1586 (2010).

条款不足以符合《美国联邦法典》第 1030 条的构成条件,如此理解会导致该条的入罪标准过于模糊。[10] 在 United States v. Nosal 案(下称“诺萨尔 I 案”)中,员工诺萨尔在离职后说服原公司的前同事帮助自己“另起炉灶”,后者利用自己的登录权限从公司计算机系统下载了一些商业数据提供给诺萨尔。法院认为,将合同法问题转化为刑法问题会导致缺乏预期与明确性,而使用条款本身很模糊而且可以被随时更改,故否定了“超越授权”的构成。[11] 合同范式和代理范式一脉相承,它们都以数据控制者的民事主观权利作为基本依据来认定授权要件。但从上述两起代表性案例可以看出,由于代理范式和合同范式会带来很大的模糊性和不确定性,上述适用标准在目前美国《计算机欺诈与滥用法》的司法实践中逐渐式微。

3. 撤销范式

所谓撤销范式,是指如果数据控制者通过一定的形式,如发送禁止令(cease and desist letter),明确表示撤销某一用户的访问权限,在此情形下用户仍然访问则构成“未经授权”。2016 年美国第九巡回上诉法院再次审理了诺萨尔案(下称“诺萨尔 II 案”),法庭修正了上述诺萨尔 I 案中的判决意见,认为既然公司已经撤销了诺萨尔的权限并与之签订一年的竞业禁止协议,那么他实际也就无权再访问公司的数据库,即使借助其他在职员工的帮助,也仍然属于“未经授权”。[12] 其后,在 Facebook v. Power Ventures 一案中,法院再次确认了这一立场。一家名为 Power Ventures 的公司,提供了一款以用户同意为前提聚合多个社交账号(如脸书、推特等)信息供用户便捷使用的软件。在该公司推广其服务的过程中,脸书公司发出了禁止令,但对方不予领会,并且绕开了脸书公司的 IP 封锁措施。法院仍然认为,单纯违反网站使用条款不构成“超越授权”,但是如果访问的授权被撤销且对方明知,则此时仍然构成“未经授权”。[13] 撤销范式试图通过一些相对明确的外部条件(如禁止令)来认定是否授权,由此数据访问者侵权或越权的主观故意也得以证明,入罪标准的客观化程度有所提升。理论上,也有学者支持并改良这一分析框架。[14] 然而,撤销范式仍然是以数据控制者所提供的合同或使用条款作为基础判断依据,虽然其明确性有所增强,但核心的入罪标准还是没有变化。因此,理论上多有批评,认为这种立场重返对《计算机欺诈与滥用法》宽泛解释的老路,取消了诺萨尔 I 案判决在免于过度犯罪化问题上取得的积极成果,[15] 甚至这种撤销的权力还潜藏着任意专断性和打压市场竞争的风险。[16]

4. 代码范式

所谓代码范式,最初由科尔(Orin Kerr)教授所倡导,其以批评合同范式为出发点,认

〔10〕 See United States v. Drew, 259 F. R. D. 449, 467 (C. D. Cal. 2009).

〔11〕 See United States v. Nosal, 676 F. 3d 854, 860, 862-864 (9th Cir. 2012).

〔12〕 See United States v. Nosal, 844 F. 3d 1024, 1040-1041 (9th Cir. 2016).

〔13〕 See Facebook, Inc. v. Power Ventures, Inc., 844 F. 3d 1058, 1068-1069 (9th Cir. 2016).

〔14〕 See Samuel Kane, Available, Granted, Revoked: A New Framework for Assessing Unauthorized Access under the Computer Fraud and Abuse Act, 87 *University of Chicago Law Review* 1437, 1462 (2020).

〔15〕 See Amber Zamora, Making Room for Big Data: Web Scraping and an Affirmative Right to Access Publicly Available Information Online, 12 *Journal of Business, Entrepreneurship and the Law* 203, 214 (2019).

〔16〕 See Annie Lee, Algorithmic Auditing and Competition under the CFAA: The Revocation Paradigm of Interpreting Access and Authorization, 33 *Berkeley Technology Law Journal* 1307, 1333-1337 (2018).

为授权的认定应当按照是否规避以代码为基础的限制来进行。[17] 在极具影响力的 hiQ Labs v. LinkedIn 案(下称“hiQ 案”)中,一家名为 hiQ 的数据分析公司,利用网络机器人爬取了职业社交平台领英用户的公开数据,领英公司不仅采取了一定的技术反爬手段,而且向 hiQ 公司发送了禁止令,但 hiQ 公司无视其爬虫协议(Robots Exclusion Standard,又称“robots 协议”),规避了 IP 封锁等措施,继续爬取数据。法院判决意见指出,《计算机欺诈与滥用法》的立法初衷是为了规制对私有领域计算机(而非对公众开放的网站)进行侵犯的行为,其援引了科尔教授偏向代码范式的观点,认为即使 hiQ 公司采取了规避 IP 封锁等措施,也仍不构成“未经授权”。[18]

代码范式对《计算机欺诈与滥用法》进行了相对严格的限缩解释,以客观层面的技术规避措施作为标准易于把握,可以在一定程度上避免司法判决中的主观性和任意性,获得了相当多学者的认同。[19] 由此立场出发,如果对代码限制或技术措施进行相对实质化的限缩解释,那么对于使用网络爬虫获取开放数据的案例,可能推导出无罪的结论。

纵览美国判例法的发展,虽然存在诸多范式之争,但关于授权问题的理解可以大体归结为两种基本标准,其一为合约权利标准,其二为技术障碍标准。在刑事责任的认定中,按照合同权利标准,对没有授权这一核心违法要素的认定主要依据数据控制者的主体权利来认定,其具有一定主观化表象;而按照技术障碍标准,授权与否则主要依托技术措施来加以认定,外在客观性更明显。通过对《计算机欺诈与滥用法》授权问题的梳理可以发现,不论是在理论界还是实务界,美国过去二十年来对数据犯罪问题的基本立场以及网络爬虫行为的罪责认定路径,逐渐从合约权利标准转向了技术障碍标准。与此同时,较之于过去宽泛解释《计算机欺诈与滥用法》的做法,尤其是对开放数据的获取而言,关于何为“未经授权”与“超越授权”的认定,目前美国法显然体现出了从严解释、逐渐限缩的趋势。[20] 由于《计算机欺诈与滥用法》相关法律条款与我国非法获取计算机信息系统数据罪具有相似性,因而对上述理论范式的辨析能够为我国的理论探讨带来启发。但是,直接将合约权利标准与技术障碍标准归结为主观和客观入罪标准,[21] 简单地移植美国的理论观点,则有将复杂问题简单化的嫌疑,对此仍需细致加以剖析。

二 基础标准的优势与不足

作为基础观察视角的合约权利标准和技术障碍标准形成于特定的时代背景和技术发

〔17〕 See Orin Kerr, Cybercrime's Scope: Interpreting Access and Authorization in Computer Misuse Statutes, 78 *New York University Law Review* 1596, 1644-1645 (2003).

〔18〕 See hiQ Labs v. LinkedIn, F. Supp. 3d 1099, 1111-1113 (N. D. Cal. Aug. 14, 2017); hiQ Labs v. LinkedIn, 938 F. 3d 985, 1001 (9th Cir. 2019).

〔19〕 See David J. Rosen, Limiting Employee Liability under the CFAA: A Code-Based Approach to Exceeds Authorized Access, 27 *Berkeley Technology Law Journal* 737, 760 (2012).

〔20〕 See Andrew Sellars, Twenty Years of Web Scraping and the Computer Fraud and Abuse Act, 24 *Boston University Journal of Science and Technology Law* 372, 393-408 (2018).

〔21〕 参见杨志琼:《美国数据犯罪的刑法规制:争议及其启示》,《中国人民大学学报》2021 年第 6 期,第 158 页。

展阶段,二者关注的侧重点有所不同。对这两项基础标准,不仅应全面分析其各自存在的优势与不足,而且也应结合本国的实际情况予以理性判断。

(一)合约权利标准的逻辑与风险

合约权利标准在美国的司法实践中曾经占据了相当重要的地位。根本原因在于,主观权利标准的形成具有一定的内在逻辑合理性。其一,合约权利标准与现有刑法构成要件的表述存在直观上的符合性。美国和中国对数据犯罪的基本规定,都没有特别地要求突破安全措施或技术障碍。在这种立法框架下,以合约等形式表现出来的相关主体授权自然而然成为基本判断依据。换言之,合约权利标准不仅没有超出构成要件的语义射程,而且接近于对构成要件的平义解释。其二,美国和中国数据犯罪的扩展性构成要件表述,也与相对宽松的合约权利标准具有内在的契合性。美国《计算机欺诈与滥用法》在"未经授权"之外规定了"超越授权",而我国《刑法》第 285 条第 2 款则采取了"其他技术手段"的兜底表述,二者都为司法适用留下了能动的解释空间,而这恰恰满足了灵活适用合约权利标准的需求。其三,合约权利标准有利于充分保护主体的数据权利。在上述诸多案例中,数据爬取行为无疑都违背了数据控制者的意愿,而且常常会为其带来较大的经济损失。数据已经成为当代社会核心的生产要素之一,出于对数据权利的积极保护,如果行为人在明显违背合约的前提下爬取数据,那么启动法律机制来予以规制也并不令人意外。

然而,美国法的判例实践表明,纯粹的合约权利标准也面临诸多值得反思的问题。其一,合约权利标准容易将民事侵权和日常行为轻易入罪,导致刑事处罚过于宽泛。越权访问数据在网络空间中相当普遍,属于一种极为常见的侵权行为,它与刑事犯罪之间的边界本身就十分模糊。而如果单纯只是依据合约权利来进行判断,在司法实践中确实可能会带来过度犯罪化的风险。其二,合约权利标准非常注重数据控制者的权利主张,但是没有对其表现形式的认定进行合理限缩,忽略了数据访问者的行为预期。在实践中,合约权利的表达方式非常多样,如使用条款、禁止令、点击同意协议等。在以海量数据为依托而建构的网络空间中,充斥着大量形形色色的数据授权合约,绝大多数用户未必会认真细致地阅读这些条款,这就容易导致数据访问者和收集者面临一种宽泛且不确定的刑事法律风险。[22] 其三,合约权利标准的带来的宽泛入罪趋势,是否符合真正的立法原意存在疑问。例如,相当多的观点认为,《计算机欺诈与滥用法》最初的立法目标主要聚焦于应对计算机黑客入侵所带来的威胁。[23] 而网络爬虫这类数据获取工具,本身具有显著的社会效用与价值,其基本适用场景是常见的互联网应用、科学研究等领域,所针对的也基本是相对开放领域(而非封闭系统内)的数据。单纯采取合约权利标准来界定数据爬取行为的刑事可罚性,违背了立法者预设的特定规制目标和打击范围。

〔22〕 See Orin Kerr, Cybercrime's Scope: Interpreting Access and Authorization in Computer Misuse Statutes, 78 *New York University Law Review* 1596, 1659 (2003).

〔23〕 See David J. Rosen, Limiting Employee Liability under the CFAA: A Code-Based Approach to Exceeds Authorized Access, 27 *Berkeley Technology Law Journal* 737, 744 (2012); Samuel Kane, Available, Granted, Revoked: A New Framework for Assessing Unauthorized Access under the Computer Fraud and Abuse Act, 87 *University of Chicago Law Review* 1437, 1442 (2020); United States v. Nosal, 676 F. 3d 854, 858 (9th Cir. 2012).

(二)技术障碍标准的优势与疑虑

在反思合约权利标准的基础上,以代码范式为代表的技术障碍标准逐渐受到理论界的青睐。较之于前者,技术障碍标准存在以下几个方面的明显优势。

其一,技术障碍标准在一定程度上避免了法律适用过程中的主观任意性倾向。按照技术障碍标准的要求,数据爬取行为是否违背授权并具有违法性以规避或突破一定的技术措施为基本依据。这种更加客观化的认定标准,不仅有利于强化法律适用的可操作性,降低司法证明难度,而且能够为数据控制者和获取者提供更为明确的行为指引。[24]

其二,技术障碍标准限缩了规制范围,避免了刑事处罚泛化。技术障碍标准将规避技术措施的客观行为作为刑事可罚性的前提,提高了入罪条件,一大批日常性的数据处理违约或侵权行为从刑事犯罪圈中剥离出来。

其三,技术障碍标准在更广阔的视野中考虑了社会利益和公共政策。在数字经济时代,数据的自由流通具有极大的社会公共价值,因此不能在刑事法律适用中无形地压缩数据访问和获取的空间。而且,互联网空间本身具有开放的属性,如果在解释论上过度偏向保护数据控制者的权利,则会在很大程度上限制数据处理者的市场竞争空间,甚至使得刑法沦为大型互联网公司打压市场竞争的工具。

虽然技术障碍标准在理论上似乎取得了越来越主导性的地位,但是仍存在一系列深层的疑问有待回答,且与中国的司法实践存在明显的裂痕。

其一,技术障碍标准在客观适用条件上并不明确。虽然技术障碍标准在一定程度上摆脱了合约权利标准的主观性和不确定性,但其并没有真正提出清楚和可靠的客观判断依据。代码是构建网络社会的技术性前提,互联网空间的几乎所有行为和机制都可以被认为以代码为基础。究竟规避或突破何种层面的技术障碍可以被认为是无权或越权访问,即使在支持代码范式的阵营中也没有一致意见。科尔教授最初提出代码限制,主要是指冒用他人认证信息(如账号秘密),以及利用代码漏洞引发程序故障从而取得更多访问权限的情形。[25] 这些情形与网络爬虫行为所采用的技术手段明显不同。一方面,利用网络爬虫抓取数据,通常并不涉及账号密码登录的问题;另一方面,网络爬虫也并非网络黑客攻击手段,除非数据抓取过于频繁,否则一般不会造成系统失灵和运行故障。面对代码理论所存在的模糊性,科尔教授实际已经对其观点作出了重大调整。他指出,用户实施诸如删除 cookie、规避 IP 封锁、绕过人机验证机制的技术规避措施,都不属于“没有授权”。[26] 而与之不同,另有基本认同代码限制路径的学者却认为,突破像“user agent”或人机验证这样的代码限制措施可以否定授权。[27]

其二,在中国法律语境下,过度犯罪化的问题是否真实存在值得怀疑。在美国的司法

---

〔24〕 See Patricia L. Bellia, A Code-Based Approach to Unauthorized Access under the Computer Fraud and Abuse Act, 84 *George Washington Law Review* 1442, 1474-1475 (2016).

〔25〕 See Orin S. Kerr, Cybercrime's Scope: Interpreting Access and Authorization in Computer Misuse Statutes, 78 *New York University Law Review* 1596, 1644-1645 (2003).

〔26〕 See Orin S. Kerr, Norms of Computer Trespass, 116 *Columbia Law Review* 1143, 1164-1170 (2016).

〔27〕 See Josh Goldfoot & Aditya Bamzai, A Trespass Framework for the Crime of Hacking, 84 *George Washington Law Review* 1477, 1487 (2016).

判例中,之所以呈现从合约权利向技术障碍标准的转变,重要的原因在于法院担心会将日常生活与工作中的轻微违法行为入罪。但是,这种担忧在我国刑法语境中没有那么迫切,因为非法获取计算机信息系统数据罪还需满足情节严重的罪量要求,对此 2011 年最高人民法院、最高人民检察院《关于办理危害计算机信息系统安全刑事案件应用法律若干问题的解释》已经作出了具体规定。此外,侵犯公民个人信息罪、破坏计算机信息系统罪这些广义上的数据犯罪罪名,同样都设置了罪量要素。罪量要素所设定的不法程度要求,能够将绝大部分轻微的无权或越权数据爬取行为予以分流出罪。

其三,过度强调突破技术障碍要求,同样可能会走向立法初衷的反面。在合理控制数据犯罪入罪边界的同时,也需平衡数据控制者的正当权利诉求。数据是许多互联网公司赖以生存的核心资产,如果失去了数据获取权限的法律保护,整个行业的经营行为预期可能落空,乃至陷入无序的恶性竞争之中。而且,对技术障碍作过于严格的解释,意味着数据控制者需要研发更加复杂的反爬手段,这不仅在技术可能性上本身就存在疑问,而且会大大提高数据控制者的企业运营成本。

综上,在数据犯罪的刑事责任问题上,美国的理论范式及其标准抉择为我们带来了重要启示。但不能忽视的是,不论是合约权利标准还是技术障碍标准都存在一定的优势与不足,而中美两国在网络数据生态和刑事司法实践层面也都有差异。在美国,由于《计算机欺诈与滥用法》过去存在被泛化适用的趋势,不仅在立法论上存在诸多批判与改良的方案,[28] 而且在司法论层面,判例也明显收缩了刑事制裁范围。以 Sandvig v. Sessions 案[29] 和 hiQ 案等代表性判例为契机,理论上甚至认为即使存在密码保护等技术障碍机制,对于涉及爬取网络数据的案例,《计算机欺诈与滥用法》可能将整体不再被适用。[30] 但是,这种整体出罪的方案,并不完全契合中国所面临的实际情况。近年来,我国利用网络爬虫恶意获取数据的行为越发频繁,数据处理和流通过程中的权利归属及其保护较为混乱,对此刑法不应完全缺位。简单笼统地从维护公共利益的角度出发,一概将该类行为排除在刑法构成要件之外,将会弱化乃至虚置数据权利的保护,最终影响数据要素市场的正常运行。

三 场景模式下的数据确权

目前在讨论爬取数据行为刑事责任的认定时,刑法学界没有对数据确权这一根本性前提予以细致分析,这会导致对问题的讨论失焦。通过网络爬虫所获取的数据通常存储

〔28〕 甚至有学者认为,最好的方案是不再直接通过像《计算机欺诈与滥用法》这样的立法来规制计算机滥用,而应通过专门的行政机构灵活设置规则来规制这些行为。See Ric Simmons, The Failure of the Computer Fraud and Abuse Act: Time to Take an Administrative Approach to Regulating Computer Crime, 84 *George Washington Law Review* 1703, 1714 (2016).

〔29〕 该案中,四名教授为了进行一项针对数据歧视性应用的研究,打算使用网络爬虫收集相关数据,但是他们担心自己的行为会因违反相关网站的使用条款而触犯《计算机欺诈与滥用法》,因而提起了诉讼。法院基于对该法的限缩解释认为,这样的行为不属于对数据无授权的访问。See Sandvig et al v. Sessions, 315 F. Supp. 3d 1, 26-27 (D. D. C. 2018).

〔30〕 See Jennie E. Christensen, The Demise of the CFAA in Data Scraping Cases, 34 *Notre Dame Journal of Law, Ethics & Public Policy* 529, 547 (2020).

于网络平台,而网络平台运营者所控制的这些数据往往并非其自身初始形成。例如,各大社交网站中的数据,绝大部分都是由用户所提供,而后由社交平台运营者进行管理。那么,数据爬取行为侵犯的是何者的何种权利,是首先需要明确的问题。平台数据权属是极富争议的问题,理论上存在数据个人所有、数据平台所有、数据个人与平台共有、数据公众所有等不同观点。[31] 但是,实际上平台管理数据的情况较为复杂,概括性进行数据确权的难度相当之大,更妥当的方案应当是在不同场景之下进行类型化分析。

(一)个人数据[32]与平台控制数据

从数据刑法保护体系的角度来看,区分个人数据和平台控制数据提供了重要的切入思路。个人数据突出对具体个人的可识别性特征,保护的法益属于人格权的具体形态,而平台控制数据则强调对数据本身作为客体的支配权限,保护法益为数据本身的私密性、完整性和可用性。[33] 因此,两种数据所对应的罪名存在根本差异,个人数据犯罪一般对应侵犯公民个人信息罪,平台控制数据犯罪则通常对应非法获取计算机信息系统数据罪(或关联性的经济与财产犯罪),二者都可能涉及网络爬虫的刑事责任问题。

其一,网络平台所控制和管理的数据中,相当一部分数据并不涉及公民个人的人格权利。例如,在引起较多关注的武汉某光公司案件中,被告通过网络爬虫和相关技术手段所获取的主要是公交车行驶实时数据,在非法获取计算机信息系统数据罪之下讨论刑事责任是妥当的。[34] 再如,单纯的房源数据,如果只涉及到房屋状况,不能识别个人,行为人通过网络爬虫从其他房产数据公司恶意抓取该类数据的案件,法院可能认定构成非法获取计算机信息系统数据罪。[35]

其二,通过网络爬虫也可能非法获取涉及公民隐私或人格的个人数据。例如,行为人利用网络爬虫从工商行政管理系统抓取大量市场主体信息(包括法定代表人姓名、电话等)予以出售,法院认定构成侵犯公民个人信息罪。[36] 又如,行为人编写爬虫软件,利用系统漏洞从上海市住房和城乡建设管理委员会网站非法获取大量从业人员信息的,同样构成本罪。[37]

其三,爬取数据的行为,可能同时侵害个人数据所承载的公民人格权和对平台控制数据的支配权限。例如,就存储在网络平台中的某些客户个人资料而言,一方面提供数据的

---

〔31〕 参见丁晓东:《数据到底属于谁?——从网络爬虫看平台数据权属与数据保护》,《华东政法大学学报》2019年第5期,第73-74页。

〔32〕 2022年12月中共中央、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》(即《数据二十条》)将数据分为公共数据、企业数据、个人数据三类。本文所称的个人数据,主要是指上述第三种数据类型中承载个人信息的数据,用于对具体个人的识别,而非个人支配或拥有的数据,所以不包括后文讨论的平台用户协议中指向的非个人信息。

〔33〕 参见王华伟:《数据刑法保护的体系考察与体系建构》,《比较法研究》2021年第5期,第145-147页。

〔34〕 参见广东省深圳市南山区人民法院(2017)粤0305刑初153号刑事判决书。

〔35〕 参见北京市朝阳区人民法院(2020)京0105刑初2594号刑事判决书。当然,如果房源信息中包含了业主姓名、电话、房屋门牌号等个人信息,则可能构成侵犯公民个人信息罪。参见最高人民检察院第140号指导性案例,柯某侵犯公民个人信息案。

〔36〕 参见辽宁省沈阳经济技术开发区人民法院(2018)辽0191刑初418号刑事判决书。

〔37〕 参见上海市徐汇区人民法院(2020)沪0104刑初731号刑事判决书。

用户并没有同意平台之外的第三方获取和使用个人信息,另一方面网络平台也在一定时间和范围内管理这些数据,这些数据甚至构成了平台运营的核心资产。针对这种情形,认定构成非法获取计算机信息系统数据罪还是侵犯公民个人信息罪,在实践中存在争议,例如,行为人发现某科技公司的电商小程序存在漏洞,通过网络爬虫爬取该公司的客户资料(包括姓名、电话、收件地址),虽然公诉机关最初以非法获取计算机信息系统数据罪起诉,但法院认定构成侵犯公民个人信息罪。〔38〕事实上,这种情形主要涉及的问题并非如何对两罪进行界分,而是如何处理两罪的竞合关系。理论上有观点认为,公民个人信息也是数据的一种,故侵犯公民个人信息罪与非法获取计算机信息系统数据罪构成特别法与普通法的法条竞合关系。〔39〕笔者认为,由于两罪的保护法益分属不同范畴,性质存在根本性差异,所以当一个数据爬取行为同时触犯两罪时,按照想象竞合从一重罪处理,可以更好地发挥其明示机能。

(二)场景模式下的权利归属

在确定数据基本类型的前提下,从数据权属的角度来看,应进一步区分不同情形分别予以讨论。平台中数据的确权规则难有非常具体的统一操作标准,但是应当主要考虑数据的初始提供者是否进行了权利授予和让渡,平台对数据的形成是否投入了相当的社会劳动成本等因素。

其一,附着于数据的权利主要由平台独有。例如,在前述武汉某光公司案件中,某米科技公司通过与公交公司合作,在公交车上安装定位器获得了公交车行驶实时数据,此时谷米公司不仅获得了公交公司的授权,而且也投入了一定成本收集、整理数据,应当认为谷米公司享有支配和处理这些数据的权利。此时,未经授权爬取数据的行为,侵犯的乃是平台的相关数据权利。

其二,附着于数据的权利由平台和用户共有。虽然理论上也存在认为平台数据由用户独有的主张,但这种理解忽视了平台的成本投入和运营需要,用户相对于平台取得了过于强势的地位,很难成为实践中的常见做法。在较多的场合,用户在平台发布数据,并未完全将数据权利全盘转移,数据权利实际由用户和平台共同享有。一方面,相对于平台而言,用户仍然享有对其上传数据的相关权利,平台更多扮演提供网络信息存储空间的网络服务提供者的角色;但另一方面,用户的这种权利又受到一定制约,尤其是不能未经许可即向第三方提供数据,以此平衡平台在激烈商业竞争中的利益诉求。例如,某手机软件用户协议中明确规定,用户通过该软件上传并发布的任何内容的知识产权归属于用户或原著作权人所有。同时,该协议又要求,对用户通过该软件上传、发布、传播的内容之全部或部分,用户授予公司在全球范围内的、免费、非独家、可再许可等权利;未经公司书面许可,用户不得自行将已发布在该软件上的相关信息内容提供给第三方。其他应用软件或平台的用户协议中也有类似的规定。此时,未经授权爬取数据,实际上同时侵犯了平台和用户

〔38〕 参见广东省深圳市南山区人民法院(2020)粤 0305 刑初 1037 号刑事判决书。

〔39〕 参见刘艳红:《网络爬虫行为的刑事规制研究——以侵犯公民个人信息犯罪为视角》,《政治与法律》2019 年第 11 期,第 24 页。

对相关数据各自享有的不同权利,但是这并不会对数据犯罪(尤其是非法获取计算机信息系统数据罪)的定性产生实质性影响,上海某品公司爬取某软件用户上传的视频一案的判决即是如此。〔40〕

其三,数据向公众一般性开放,则应视为取得了数据权利人的推定同意。网络平台管理的某些数据,出于权利人的主观意愿或者某些法定理由,可能处于无需特别授权的一般开放状态,而非仅仅针对特定目标群体开放,此时即使利用网络爬虫加以获取,也应原则上否定行为的构成要件符合性。例如,为了落实审判公开原则、促进司法公正,中国裁判文书网依据《最高人民法院关于人民法院在互联网公布裁判文书的规定》向社会大众一般性公开法律文书,爬取该类数据应当视为获得了推定授权。此外,在一些开放性的社交网络中,用户所展示的数据是否具有一般开放性的特征仍然存在较大争议,这很大程度上取决于社交网络运营者与用户之间通过用户协议如何进行具体约定,数据权利人是否具有向社会公众开放数据的意愿,以及这些开放数据是否已经构成了公共信息的一部分。在 hiQ 案中,法院之所以判决领英败诉、否定数据爬取者刑事责任,根本性的考虑即在于此。〔41〕 在 hiQ 案最新的重审判决中,第九巡回法院更是明确指出,《计算机欺诈与滥用法》适用的基本前提应当是区分一般可访问的信息和需要授权的信息,而本案中 hiQ 公司抓取的数据并不属于领英公司,而用户也没有明确表达对其公开发布信息的隐私期望。〔42〕 当然,需要注意的是,并非所有社交平台中的数据都一概处于一般性开放状态,其权属状态仍然取决于具体语境中权利人的主观意愿及其与平台之间的约定。

四 合约权利与技术障碍的双重确证

在数据确权的基础上,在数据犯罪构成要件所划定的语义射程范围内,何种数据爬取行为具有实质的违法性,应当结合合约权利和技术障碍双重标准来确定,两者并非相互排斥的择一关系,而是构成逐层递进的刑事不法补强逻辑。

(一)违背合约权利奠定基础不法

1. 作为可罚性基础的合约权利

如果数据完全处于封闭和保密状态,行为人借助技术手段,恶意攻破对方安全措施或秘密侵入对方计算机信息系统非法爬取数据,认定构成相关犯罪并无太大争议。网络爬虫案件背后真正复杂的问题在于,由于数据本身已经处在相对开放的状态,何以能够证立数据爬取行为的刑事违法性。本文认为,对数据权利人合约权利的侵害,奠定了网络爬虫刑事责任的基础不法。在网络爬虫相关的刑事案件中,涉及的数据权利几乎都属于个体法益。个体法益的刑法保护意味着,应当充分尊重权利人对法益的处分权限和处分方式。网络平台中的数据控制者,恰恰是希望有条件地处分自己的数据权利:用户可以自由地访

〔40〕 参见北京市海淀区人民法院(2017)京 0108 刑初 2384 号刑事判决书。

〔41〕 See hiQ Labs v. LinkedIn, 938 F. 3d 985, 1003-1004 (9th Cir. 2019).

〔42〕 See hiQ Labs v. LinkedIn, 31 F. 4th 1180-1183, 1190-1191, 1199-1200 (9th Cir. 2022).

问和获取开放数据,但是不得使用网络爬虫或相关软件程序大规模批量获取。恶意使用网络爬虫获取数据,实际是通过违反合约的方式,侵害数据权利人的数据处分权限。如有学者所言,授权与否的判断,主要不是看用户做了什么,而是数据持有者允许和同意用户做什么。〔43〕 爬取数据行为的实质违法性,本质上根源于被害人同意的判断,〔44〕而合约恰恰是同意的主要外在表现形式。

合约权利的判断可能存在主观性、不确定性的缺陷,这一点在美国刑事司法判例中已有充分讨论。但是,这些问题并不能改变合约权利侵害对刑事违法性认定的基础性意义。就网络爬虫的刑事责任而言,理论上对合约违反的刑法评价存在较大分歧。有观点认为,只要不满足技术层面是否具有侵入性、是否遵守合同约定、是否具有目的正当性三个条件中的任何一个便可认定数据爬取行为的非法性。〔45〕 但是,这样的理解没有对单纯合约权利路径的刑法扩张趋势给予合理限定,这可能带来轻易入罪的风险。对合约授权的违反是奠定行为刑事违法性的必要不充分条件,它构成后续突破技术障碍判断的前提。

2. 作为核心认定标准的爬虫协议

鉴于合约权利标准存在的不足,应当对合约形式及其内容进行限定解释。数据控制者在平台所设定的合约形式多种多样,但并非所有涉及数据的合约都与爬取数据行为的刑事违法性相关联。例如,各大社交平台或电商平台都会提示一定的使用条款,告知用户使用服务的相关注意事项,其中也包括一些授权条件。但是,并非违背了使用条款的行为一概都应被评价为数据犯罪意义上的未经授权或超越授权。比如,用户违反了使用条款中的实名制要求,并不必然导致访问、获取数据行为具有刑事可罚性,因为实名制要求主要服务于用户管理等政策性目标,与数据授权问题不具有规范目标上的一致性。准确地说,只有当合约条款的内容直接指向数据访问和获取的授权,并且行为人明确违背这一合约,才可能涉及到刑事责任问题的讨论。

正是在这一意义上,爬虫协议的刑法效力值得重点辨析。所谓爬虫协议,通常是一个存放于相关网站中的纯文本格式文件(文件后缀名为“.txt”),它在网络爬虫程序访问时告知对方哪些内容不能被爬取。一方面,它虽以文本文件的状态存在,但是并没有技术层面的强制性效力,无法构成技术障碍;另一方面,爬虫协议虽然名为协议,但是并非双方合意签署,而是数据访问、获取授权范围的单方声明,可以理解为一种广义的合约(或条约)。对于爬虫协议的违反,能否认定数据爬取行为具有刑事违法性,理论上存在截然对立的观点。肯定性的意见主要认为,爬虫协议在行业内乃至司法判决中已经得到了广泛认可。〔46〕 而反对的观点则认为,爬虫协议只是一种君子协定因而约束力很弱,而且将其作为违法性判断依据容易将犯罪判断权委以数据网站。〔47〕 本文认为,在使用网络爬虫恶

〔43〕 See James Grimmelmann, Consenting to Computer Use, 84 *George Washington Law Review* 1500, 1500-1501 (2016).

〔44〕 类似观点,参见林维:《数据爬取行为的刑事司法认定》,《人民检察》2020 年第 4 期,第 39 页。

〔45〕 参见苏青:《网络爬虫的演变及其合法性限定》,《比较法研究》2021 年第 3 期,第 97 页。

〔46〕 参见刘艳红:《网络爬虫行为的刑事规制研究——以侵犯公民个人信息犯罪为视角》,《政治与法律》2019 年第 11 期,第 22 页。

〔47〕 参见薛美琴:《网络爬虫刑法规制的边界》,载杨明主编《网络法律评论》2020 年第 23 卷,中信出版社 2021 年版,第 236-237 页。

意获取数据的案件中,爬虫协议应当成为合约授权的核心认定标准。

其一,设置爬虫协议的做法已经成为互联网领域和数据处理领域的行业惯例。网络空间的规则之治仍处在形成与发展过程中,对于通过网络爬虫获取数据行为责任边界这样的新型问题,相应行为规范的成熟度与效力层级不宜过度苛求。尽管爬虫协议的效力和地位没有在成文法中被明确,但事实上早已成为业界关于获取数据授权边界的基本共识。早在2012年,多家互联网企业就共同发起并由中国互联网协会发布了《互联网搜索引擎服务自律公约》,其中第6条明确规定,遵循国际通行的行业惯例与商业规则,遵守机器人协议。这充分说明,设置爬虫协议已经成为具有广泛认可度、具有准规范性的行业惯常做法。

其二,爬虫协议的内容直指数据访问和获取的授权问题。合约权利标准之所以在美国的司法判例与理论文献中饱受诟病,很重要的原因之一在于,没有对合约的范围进行严格限定。爬虫协议的核心内容在于,明确对外声明,具体何种范围内的数据允许或不允许被爬虫抓取。可以认为,爬虫协议是网络平台经营者针对爬虫访问和获取数据授权范围所作出的最直接和清楚的意思表示。除此之外,在纷繁复杂的数据流通环节中,几乎没有更为权威和可靠的数据授权依据。因此,不论是在国外还是国内的诸多典型案例中,[48]爬虫协议对数据授权的效力在司法判例中已经反复得到了确认。即使是在百度公司与奇虎360搜索公司之间的不正当竞争纠纷案中,虽然法院始终认定百度公司设定爬虫协议限制360搜索引擎抓取内容的行为构成不正当竞争,[49]但是法院判决也并未否定爬虫协议本身的正当性。

其三,通过爬虫协议所形成的选择性合约授权,是当下许多互联网企业核心运营模式得以存续的重要前提。在信息和数据服务领域,许多互联网企业对于数据获取权限常常面临两难困境。一方面,在所谓"流量为王"的网络时代,互联网企业需要通过向用户开放数据获取路径,才能逐步扩展市场,奠定良好用户基础;另一方面,开放数据如果轻易被其他竞争者在短时间内批量获取,那么互联网企业又面临市场优势迅速瓦解的风险。因此,选择性地开放数据,尤其是通过爬虫协议对网络爬虫这类数据批量抓取工具予以保留,在很大情况下成为不得已的选择。

(二)突破技术障碍确证刑事不法

1. 作为限制处罚条件的技术障碍

按照我国刑法数据犯罪的规定,技术障碍的突破并没有在立法表述上被强调,但是在合约权利侵害认定的基础上,技术障碍对限定网络爬虫刑事处罚范围发挥了重要的作用。

第一,技术障碍的突破外化了合约权利的违反,使其能够以更为清晰和客观的标准予以把握。在网络在线交往的语境下,平台运营者为了规避可能的法律风险,往往会制定诸多非常复杂的合约条款让用户签署。形形色色的条款让人眼花缭乱,用户往往没有时间

〔48〕 参见张金平:《有关爬虫协议的国外案例评析》,《电子知识产权》2012年第12期,第81页以下;杨华权、曲三强:《论爬虫协议的法律性质》,《法律适用》2013年第4期,第33页。

〔49〕 参见北京市第一中级人民法院(2013)一中民初字第2668号民事判决书;北京市高级人民法院(2013)高民初字第3755号民事判决书;北京市高级人民法院(2017)京民终第487号民事判决书。

和耐心深究这些合约,因而人们常常并不真正清楚访问和获取数据的权利边界何在,由此形成了不确定性的法律风险。在爬虫协议基础上所设立的客观技术障碍,能够在很大程度上消除以上的不确定性。一方面,技术障碍的设置,非常清楚、具体地表达了平台对数据访问和获取的授权限制意思;另一方面,如果用户想要突破平台设置的技术障碍,需要针对性地采取技术手段加以攻克,在这一过程中原则上不可能再对平台中的数据权属和相关法律风险存有疑问。

第二,技术障碍的突破在一定程度上提升了行为整体的不法程度,为网络爬虫的刑事责任隐性地设置了一定入罪门槛,防止处罚泛化。为了维护自身核心的数据权利不被侵犯,平台运营者不得不投入较大成本设置技术保护措施,并且定期更新维护,构筑数据安全壁垒。而行为人突破乃至破坏技术保护措施,不仅会为平台的数据安全维护带来经济损失,而且对数据权利形成了更高程度的威胁。与此同时,蓄意突破平台的技术障碍,严重违背权利人的意愿抓取数据,这也体现出了行为人更高的主观不法。通过技术障碍标准的设置,能够建立起更为突出的刑事违法性准入门槛,防止出现网络空间和数据流通环节的过罪化风险。

第三,通过增加突破技术障碍的客观条件,能够更好界分刑事责任与民事纠纷。随着互联网经济的兴起,网络爬虫所涉及的民事纠纷早已有之,且不在少数。虽然民事侵权与刑事责任并不相互排斥,但是考虑到诸多数据商业模式仍然处在探索阶段,数据界权问题也仍然存在很大争议,数据要素市场的建立需要更多激励措施而非过严威慑,这种情形下宜在民事经济纠纷和刑事犯罪之间拉开一定距离。

2. 作为主要技术障碍的反爬措施

尽管引入美国技术障碍标准的合理性得到了越来越多的支持,但是技术障碍应当具有何种形式与内容,仍然存在相当大的争议,这恰恰是该标准的不足之处所在。如上文所言,美国法语境中的代码范式,并没有真正清楚地说明,何种意义上的代码限制与刑事违法性直接相关。在我国目前主张技术障碍标准的学说中,同样存在诸多内在的对立,核心的问题主要在于典型的反爬虫措施是否构成一种具有刑法评价意义的技术障碍。肯定性的观点认为,故意避开或强行突破 IP 限制、验证码等反爬虫措施的网络爬虫需要承担刑事责任;〔50〕而否定性的意见则几乎全盘接受了科尔教授的观点,认为绕开 IP 地址和验证码等反爬虫机制,并不属于突破安全保护措施,不应构成犯罪。〔51〕后一种观点实际上从技术障碍标准转向了身份认证标准,其主张 IP 地址封锁并非真正的障碍,因为用户即使在正常使用时也可能会周期性地更换 IP 地址;至于服务于人机测试的验证码,应当理解为降低用户访问速度的机制,而非对授权的否认。〔52〕但是,否定性的理解至少并不符合中国当下数据服务市场的实际情况,不利于数据要素市场的长期稳定发展。

首先,滥用数据抓取工具强大的技术能力,可能背离数据开放的初始目标。按照上述

〔50〕 参见杨志琼:《数据时代网络爬虫的刑法规制》,《比较法研究》2020 年第 4 期,第 195 页。

〔51〕 参见孙禹:《强行爬取公开数据构成犯罪吗》,《国家检察官学院学报》2021 年第 6 期,第 133-134 页。

〔52〕 See Orin S. Kerr, Norms of Computer Trespass, 116 *Columbia Law Review* 1143, 1168, 1170 (2016).

论者的理解,网络爬虫的数据访问与收集,与自然人并无本质区别,只是快慢程度有所不同。一种典型的观点认为,网络爬虫不过就是自动化的用户点击而已。〔53〕因此,在涉及数据犯罪刑事责任的核心授权问题上,网络爬虫与自然人也不应存在差异。诚然,不论自动运行的网络爬虫还是手动操作的自然人,都是按照既定路径访问和获取数据。但是,二者在处理数据的速度和规模方面存在极大差异,量变引起质变,因而不能简单将其等同看待。对一般的自然人而言,访问和获取数据的数量和规模始终是有限的,这符合数据控制者有条件开放数据的初始目标。相反,网络爬虫在一定时段内持续、高速、海量地抓取数据,可以使得数据权利人几乎整体上丧失对数据的独占性支配,后续的经营活动将遭受严重影响。

其次,上述观点借助模糊不清的身份认证标准,实际设置了过高的技术障碍要求。在网络空间,按照数据权利人的意愿,对数据访问和获取的授权根据身份进行识别,这本身是一种合理的做法。但是,在有条件开放数据的平台运营模式中,数据权利人所追求的身份识别目标恰恰是区分机器(网络爬虫)和人。具体而言,反爬虫措施就是这种身份识别追求的外在表现。上述观点对网络爬虫的技术特性轻描淡写,弱化反爬虫机制的刑法意义,实际上否定了人机身份识别在数据授权过程中的价值。照此逻辑,只有破坏更为严格的数据底层安全防护机制,才可能构成刑事犯罪。如果本已耗时耗力的反爬虫措施仍然不能维护自身对数据的合理支配权限,互联网企业可能不再敢于对外开放数据,转而对数据筑起高高的安全技术壁垒,这将导致数据流通空间和自由度迅速压缩。

最后,更深层的问题在于,以上论者并没有全面审视当下的数据运营模式,忽视了对企业数据权利、多元商业模式、用户信息自由等因素妥当地进行利益平衡。上述观点认为,权利人不能在向世界开放数据的同时,又仅仅通过表达意图的方式否定特定用户的数据权限;即使否定刑事责任的构成,公司的商业策略也不会受影响。〔54〕但是,数据的访问和获取权限,并非总是要么完全无条件开放,要么完全封闭。不同于传统经济,网络经济更注重免费性,以此获得人们的注意力这一珍稀资源,从而实现长久经营。〔55〕但与此同时,如果潜在的竞争对手采取技术手段批量“薅羊毛”,那么这种经营策略就将迅速溃败。因此,对数据有所保留的开放,恰恰是数据平台企业在激烈市场竞争中能够得以存续和发展的重要策略,在一定条件下也是反对不正当竞争的自我保护手段,当然,这些措施有时也可能会异化为进行不正当竞争的手段。平台企业在收集处理数据的过程中,投入了大量时间与金钱成本,如果否认其选择性开放的数据处分权利,将会在很大程度上挫伤其继续提供优质开放数据服务的积极性。

总而言之,对于网络爬虫行为刑事责任的认定,在数据确权和厘定法益归属的基础上,应当通过限定解释的方法,将合约权利标准与技术障碍标准加以结合,二者互相补足,缺一不可。一方面,对合约权利的违反,奠定了网络爬虫行为的基础不法,否则不具有法

〔53〕 参见薛美琴:《网络爬虫刑法规制的边界》,载杨明主编《网络法律评论》2020 年第 23 卷,中信出版社 2021 年版,第 244 页。

〔54〕 See Orin S. Kerr, Norms of Computer Trespass, 116 *Columbia Law Review* 1143, 1169-1170 (2016).

〔55〕 参见[美]凯文·凯利著:《网络经济的十种策略》,萧华敬、任平译,广州出版社 2000 年版,第 71 页以下。

律规制的必要性。而且,只有像爬虫协议这样直接指向数据访问和获取授权的合约才能符合要求,这为网络爬虫行为的罪责认定筑起了第一道相对明确的门槛。另一方面,鉴于合约权利标准具有主观化、相对化的不足,仅此还不足以清楚和合理地界分民事责任、行政责任与刑事责任。只有当网络爬虫行为对数据权利的侵犯达到了较为严重的程度,才需要刑法的介入。突破以反爬虫措施为代表的技术障碍,不仅印证了对合约权利的违反,而且在客观和主观上都提升了行为的整体不法程度,使其与日常失范行为、轻微侵权行为得以区分。同时我国刑法数据犯罪的构成要件中已经设置了罪量要求,这些因素的递进叠加判断可以实现在第一次法(民法、行政法等)和第二次法(刑法)之间划定相对合理的距离。但是,对技术障碍标准的具体认定要求,也不能像美国部分文献那样对其过度拔高,而完全忽视对平台运营者的利益平衡。根据本文所主张的"合约权利+技术障碍"二元标准,在前述理论上广受热议的上海晟品公司案和武汉元光公司案中,网络爬虫行为同时具备了对数据合约权利的违反和对反爬机制的突破,且因此给权利人带来较大财产损失,认定构成非法获取计算机信息系统数据罪是合理的。

五 结 语

网络爬虫越来越多地被运用于生产生活多个领域,但是数据犯罪的刑事风险也伴随而来。认定网络爬虫可能涉及的刑事责任的核心难题在于授权与否的判断。不同的数据犯罪规制结构,不仅直接影响爬取数据行为刑事责任的边界,而且也限定了授权问题的认定路径。中美两国数据犯罪的构成要件都采取了相对宽泛的表述,对二者授权认定范式进行比较考察具有重要价值。在美国《计算机欺诈与滥用法》的适用过程中,关于数据犯罪中的"未经授权"或"超越授权",司法判例和理论研究对此发展出来代理范式、合同范式、撤销范式、代码范式等不同认定路径。在诸多范式背后,主要是合约权利标准和技术障碍标准的对立。这两种基础标准各自都有自身的优势和不足,在中国法的背景下可以对此进行借鉴和参考,但两者不应理解为相互对立的关系。关于网络爬虫的刑事责任,应当首先通过场景式、类型化的思路,进行数据确权分析。平台中数据权利的归属争议极大,目前理论上对网络爬虫行为罪责的认定,恰恰缺少了对这一关键性问题的讨论。具体而言,应对个人数据与平台控制数据进行区分,并结合不同情形判断数据权利在个人、平台、公众之间的归属。在此前提下,对合约权利的违背奠定了网络爬虫行为的基础不法。基于爬虫协议的特殊性质,其应当成为合约授权的核心认定标准。以此为基础,突破技术障碍可以进一步提升和确证刑事不法,以此实现网络爬虫刑事责任认定的相对明确和谦抑。通过合约权利和技术障碍的二元结合标准,可以将严重侵犯数据权利的爬虫行为,与日常失范行为、轻微侵权行为区分开来。在此认定过程中,应当重视数据企业和平台的运营模式,合理平衡多方主体的利益,思考真正符合中国国情的理论解决方案。

[本文为作者主持的 2020 年度教育部人文社会科学研究青年基金项目"刑法罪量要素的理论建构与实务应用研究"(20YJC820045)的研究成果。]

[**Abstract**] Web crawlers are more and more commonly used in many fields of production and life, but the criminal legal risks of data crimes also come with them. The core issue in the determination of the boundary of criminal liability of web crawlers is the authorization for obtaining data. Different regulatory structures of data crimes not only directly affect the boundaries of criminal liability for data crawling, but also limit the path to the determination of the authorization issue. The data crime provisions in the criminal laws of China and the United States have both adopted a relatively broad expression, so it is of great value to analyze the paradigms for determining the authorization issue in the United States. Regarding the determination of "no authorization" or "exceeding authorization" in data crimes, there exist different paradigms in American jurisprudence, such as the agency paradigm, the contract paradigm, the revocation paradigm and the code paradigm. The criteria of contractual rights and those of technical barriers represented by the above paradigms have their own advantages and disadvantages. The two sets of criteria are not mutually exclusive but should be in a complementary relationship with each other. To determine the criminal liability of web crawlers, we should first analyze the data ownership with scenario-based and typological thinking. However, the ownership of data rights in network platforms is highly controversial and discussion about this crucial issue is lacking in the current analysis of the criminal liability of web crawlers. Specifically, personal data should be distinguished from platform control data, and the data rights of individuals, platforms, and the public should be ascertained in light of different situations. On this basis, the violation of contractual rights lays the foundation for the illegality of data crawling. The form of the contract should be limited and the Robots Exclusion Protocol plays an important role in this respect. The breakthrough of technical barriers further enhances and confirms criminal illegality, thereby limiting the excessive expansion of punishment. The identification of technical barriers should not be too strict. Anti-crawler measures can be classified into this category and the balance between multiple interests such as enterprise data rights and platform operation mode should be taken into proper consideration. Through the combined criteria of contractual rights and technical barriers, data crawling behaviors that seriously violate data rights can be distinguished from daily misbehavior and minor infringements. In this way, data rights are fully protected and the last-resort nature of criminal law is maintained. The determination of the criminal liability of web crawlers requires us not only to carry out comparative research by taking relevant paradigms of data authorization as reference but also to carefully observe the reality of the Internet economy and the internal logic of the platform business model in the context of localization. Only in this way can we propose theoretical solutions to the problems relating to controversial data processing behaviors that truly suit China's national conditions.

(责任编辑:贾　元)