

不当数据抓取现有规制路径的反思和创新

——以教育领域数据抓取和使用为例

任玉欣

华东政法大学

摘要: 随着教育数字化转型深入,教育数据成为驱动教学创新、评价改革与管理决策的核心要素。然而,针对教育数据的抓取、收集与滥用行为日益频发,这不仅危及个人隐私与数据安全,更有可能扰乱教育评价公平性、侵蚀学术诚信基础,甚至扭曲教育生态。现有研究多集中于商业领域数据,且仅从静态的法律责任角度审视,难以应对教育数据主体多元——包括学生、教师、学校,以及利益复杂的特有情境。本文主张,应超越事后问责的传统法律思维,构建针对教育数据治理的立法、司法与行政监管三元动态互动框架。相比于单一的立法完善或司法裁判视角,甚至立法-司法的二元互动机制。该框架强调,立法确立规则基线与基本原则,司法通过个案裁决厘清行为边界、形成规则补充,而行政监管则发挥其专业性与主动性,从事前合规指引、事中动态监测到事后快速反应,进行全程介入。三元主体在互动中协同演进,共同构成一个“事前预防—事中干预—事后救济”的完整治理链条,从而实现从静态法律条文的适用到动态系统治理的范式转型,以平衡教育数据流通利用、技术创新与教育主权、学生发展权益保护之间的多元价值。

关键词: 数据抓取;教育数据;三元治理链条;动态治理;教育数据主权

DOI: 10.65976/3080-0374.2026.06.027

引言

教育数字化转型在重塑教学模式、赋能教育治理的同时,生成了海量、敏感且具有高价值的“教育数据”,包括学生个人信息、学习行为轨迹、教学资源、学术成果及管理信息等。这些数据不仅是宝贵的教育研究资源,也成为某些不法行为觊觎的目标。实践中,针对教育数据的抓取行为已呈现出多样化形态:在学术场景,存在利用爬虫技术批量抓取学术论文、课程资源以进行论文代写、作业抄袭乃至运营“论文工厂”的现象;在教学与管理场景,未经授权抓取在线教育平台核心课程内容、学生课堂互动数据用于不正当竞争,或非法爬取招生、就业信息进行倒卖的事件时有发生;在评价场景,通过技术手段抓取并干扰评教、评奖数据,或在教育类应用中“刷分”“刷榜”,破坏评价公正性与市场秩序。

更严峻的是,教育领域已成为数据安全事件的高发区。从“学习通”超1.7亿条信息泄露,到多地高校、中小学因系统漏洞导致数千乃至百万量级的学生、家长、教师信息在暗网被售卖,一系列案例表明,教育数据的非法获取与滥用已从潜在风险演变为现实威胁,其危害远超商业领域,直接关涉未成年人保护、教育公平底线与学术伦理根基。

然而,面对这一紧迫议题,现有治理体系存在显

著短板。当前规制讨论多集中于《中华人民共和国反不正当竞争法》《中华人民共和国网络安全法》等一般性法律在商业领域的适用,缺乏针对教育数据特殊属性的专门考量。

一、教育数据不当抓取法律责任

数据抓取行为,是指利用自动化算法程序,按照指定规则自动提取并下载所需的网页数据,进而形成互联网网页镜像备份数据。抓取技术本质上是中立的。

但是随着数据抓取技术广泛应用,现实生活中非法抓取行为频发。数据抓取纠纷可能带来的法律责任有三个类型,分别是刑事责任、行政责任、民事责任。

(一) 刑事责任

在刑事责任中可能触犯的罪名是非法获取计算机信息系统数据罪、侵犯公民个人信息罪、侵犯商业秘密罪。对应的行为包括突破技术防护、获取数据、非法获取大量个人信息窃取商业秘密等。

(二) 行政责任

行政责任中可能违反的法律是《中华人民共和国网络安全法》《中华人民共和国数据安全法》《网络数据安全管理条例》《中华人民共和国反不正当竞争法》。对应的行为包括违反数据安全管理制度、干扰网络服务正常运行、进行虚假宣传等。

一个非常典型的数据抓取而承担行政责任的案例是“全国首例电商平台商品数据‘一键搬家’行政处罚案”。王某明实际控制的某枫公司、某陶公司开发并向其他电商平台经营者有偿提供四款“搬家软件”。这些软件利用爬虫技术，未经授权绕开淘宝和天猫平台的验证机制及反爬措施，抓取两平台海量商品数据，并搬运至其他电商平台开设“无货源店铺”。经查，其数据库服务器中“taobao”表数据多达1.63亿条。镇江市市场监督管理局依据《中华人民共和国反不正当竞争法》第二十四条的规定，对涉案公司作出责令停止违法行为，并处罚款的行政处罚。该案也被称为全国首例针对此类行为的行政处罚决定。

（三）民事责任

民事责任可能违反的法律是《中华人民共和国反不正当竞争法》《中华人民共和国民法典》《中华人民共和国著作权法》。对应的行为包括爬取数据实质性替代服务构成不正当竞争、违反合同约定、侵犯他人著作权。

二、立法供给的滞后

当前，我国涉及数据的法律法规，如《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，为数据治理提供了顶层框架，但多为普适性规定，难以精准覆盖教育数据的特殊性。教育数据具有鲜明的强伦理性、公共产品属性与发展性特征，其核心主体多为未成年人，关涉人格尊严与发展权。现有立法未能充分回应这些特性，导致规制针对性不足。

具体而言，在国家层面，缺乏针对教育数据，特别是其抓取、利用行为的专门立法。教育领域的数据规定散见于《未成年人学校保护规定》等规章中，多以原则性宣示为主，如“保护学生个人信息”，但对“抓取”这一特定行为模式缺乏禁止性规定与罚则。在地方层面，尽管部分省份出台了数据条例，但涉及教育数据的规定同样笼统。

未来的立法完善，应致力于填补这一空白。在《中华人民共和国学前教育法》《终身学习法》等教育专门法律的制定或修订中，应设立教育数据管理的专章或条款。更为迫切的是，应推动制定部门规章层级的《教育数据管理办法》，明确：（1）教育数据的分类分级标准，如区分公开教育资源、脱敏研究数据、核心教学数据、敏感个人信息；（2）数据抓取的正面清单，明确禁止抓取未成年人生物识别、精准轨迹等敏感数据，规范对公开教育资源的合理使用；（3）各主体的

安全责任，厘清学校、平台、第三方服务商在委托处理、共同处理等场景下的责任边界。通过立法确立教育数据治理的公益导向与底线规则，为后续的司法裁判与行政监管提供清晰的规范基础。

三、数据不当抓取责任承担界限模糊

（一）数据抓取产业链责任分配比例标准不一

数据抓取不正当竞争行为呈现产业链特征。产业链环境上存在多个主体，如数据抓取者、数据抓取使用者。总体来说数据抓取纠纷中，关于各方谁应当承担、应当承担多少比例的责任目前法律没有明确的规定。比如企业将技术外包他人，负责数据抓取的技术提供者是否应当承担。通过统计数据，发现已经颁布数据条例的6个省市（江苏、上海、湖北、湖南、宁夏、甘肃）中，仅有《江苏省数据条例》第63条明确规定了“多个数据处理者存在时各承担相应安全责任”。其他省份，《上海市数据条例》第78条，仅规定“数据处理者是数据安全主体”，《湖北省数据条例》第50条仅规定“数据处理者是数据安全主体，数据处理者应当建立健全全流程，数据安全管理制度，明确数据安全负责人和管理机构，落实数据安全保护责任”。上述规定的共性在于都是以原则性规定为主，缺乏建立健全管理制度，未细化委托管理、共同处理等具体场景下的责任比例划分标准，缺乏可操作性。

（二）数据权属不清

数据爬取行为的正当性没有统一的裁判标准。有学者认为，倘若能够明确数据权属数据爬取行为的正当性判定问题也就迎刃而解。数据权属问题学界至今意见纷纭，司法判决亦无定论，

对于上述学者观点，笔者认为此种观点过于片面。数据权属是解决争议的一个关键因素但其观点将复杂问题过于简单化了。学者的逻辑可以认为只要确定了数据归谁，那么别人抓取就是不正当，反之则是正当。诚然，数据归谁所有或者控制，确实是解决谁有资格主张权益的前提。但是在特定场景下去衡量抓取行为是否正当，需要进行利益的综合权衡考量的因素远不止于权属。有权属不等于可以绝对禁止他人接触。其次，当今数据形态多样，权属问题在实践中非常难界定。比如一组数据可能同时包含个人信息，数据集合和衍生数据产品。那么它的权属并非单一、排他的所有权，而是一个包含个人信息权益数据、财产性权益、知识产权的综合体。

通过对司法案例如微博诉脉脉案、抖音诉刷宝案裁判逻辑的拆解分析，法院的裁判逻辑从纯粹的权利

保护思维逐渐转向秩序维护思维。在数据分类基础上回归反不正当竞争法行为规制法的属性,关注爬取方式、数据使用对市场竞争秩序可能产生的影响。在行为正当性的权衡天平上,权属仅仅只是权衡要素之一。另一端还存在着鼓励数据流通、促进技术创新、保护消费者福利、维护公共利益等对立因素。

(三) 智能规制理论

我国学者在伊恩·艾尔斯和约翰·布雷斯回应性监管理论的基础上,针对数据抓取纠纷当下存在的规范供给断层、事实认定梗阻与责任归属困境提出智能规制理论。这一理论符合当今数据动态多元非排他的特性。不再执着于界定一个僵化的排他的数据所有权更具灵活性和适应性。除此之外,面对技术性极强的数据抓取行为,这一理论可以实现“以技术管技术”,比如利用大数据监测异常流量,同时弥补人工取证的缺陷。

四、调研分析:咸鱼平台视角下数据抓取服务市场的生态观察

笔者收集了闲鱼上32家提供数据抓取服务的商家的广告宣传,统计其销售时使用的诸如“免责声明”“涉嫌违法违规的技术暗词”“可抓取数据范围”“售后服务”等特定话语,这些提供数据抓取技术的商家的商业模式是一种典型的“技术分包”模式。上游是目标网站或APP,下游是各类有数据需求的客户、学生、研究者、小商家等。他们利用信息差和技术差,将复杂的爬虫开发能力以极低的价格商品化、标准化出售。

(一) 咸鱼数据抓取服务市场分析

从业务范围来看,这些提供数据抓取服务的闲鱼商家普遍严格限制抓取数据条数,这既是控制成本和风险的方式,也暗示了其规避法律风险的意图——小批量抓取不易触发平台的反爬机制和法律追责。他们抓取的数据往往涵盖国内外各种网页和app,其中,社交媒体的用户评论、私信,以及通过APP接口才能获取的数据,很多都属于受法律严格保护的个人信息范畴。根据《民法典》第一千零三十四条“个人信息的定义”,个人信息是以电子或者其他方式记录的能够单独或者与其他信息结合识别特定自然人的各种信息,包括自然人的姓名、出生日期、身份证件号码、生物识别信息、住址、电话号码、电子邮箱、健康信息、行踪信息等。《中华人民共和国个人信息保护法》第四条第一款规定个人信息是以电子或者其他方式记录的与已识别或者可识别的自然人有关的各种信息,不包括匿名化处理后的信息。抓取这类数据极易践踏法律红线。尽管如此,仍有15.6%的商家声明一切皆

可爬(他们声明“没有我们爬不到的数据”),表明了他们尚未意识到抓取数据的法律风险性,或对其抱有防之任之的侥幸心理,对于这类信息的保护意识薄弱,极易造成法律风险。从使用技术来看,根据上面的表格易得有近半数的商家提到了可使用“规避反爬机制”“JS逆向”“绕过滑块验证”“高并发抓取”等技术,这明27确表明其服务内容不是简单的公开信息采集,而是会突破网站采取的技术保护措施,对目标网站服务器造成巨大压力。这在法律上极易构成“破坏计算机信息系统”或“不正当竞争”。从宣传方式来看,有21.8%的商家给贴上了学生价的标签,并有12.5%的商家自称自己的本职工作是爬虫工程师,他们利用学生、工程师身份进行信用背书,并保证自己技术成熟,有31.2%的商家提到完整售后,保证质量,合理化其低价策略,降低消费者戒备心理。

(二) 风险转嫁与教育缺口

然而,这一灰色市场的蓬勃,与其下游需求的支撑密不可分。本调研显示,有21.8%的商家特意标注“学生价”,其目标客户群体不言而喻。对于在校大学生而言,他们可能因课程设计、论文研究、商业尝试或单纯的技术好奇,成为此类数据抓取服务的购买者与使用者。但必须清醒认识到,“能用技术实现”绝不等于“法律允许”或“合乎伦理”。上述商家所兜售的“规避反爬机制”“JS逆向”等服务,其本质是教唆、协助用户实施一系列可能构成违约、侵权乃至违法犯罪的行为。学生一旦购买并使用此类服务,不仅其获取的数据及衍生成果如作业、论文、商业项目,在法律和学术伦理上存在根本瑕疵,更可能使自己直接卷入法律纠纷,面临平台封禁、民事索赔乃至行政处罚的风险,对个人信用与职业生涯造成不可逆的负面影响。

更重要的是,部分掌握技术的学生可能从“消费者”转变为“供给方”,出于兼职或兴趣,模仿此类模式在闲鱼等平台承接抓取业务。这无疑是从“风险关联方”滑向“风险制造者”,将面临更直接的法律制裁。因此,无论是作为消费者还是潜在从业者,大学生群体都必须打破“技术中立”的迷思,深刻理解:任何绕开或破坏技术保护措施的数据抓取行为,都已踏入法律的红线。高等教育的目标是培养德才兼备的负责任人才,而非精通“黑客”技术的法盲。在数字时代,技术能力必须与法律素养、伦理意识同步提升。学校与社会应加强引导,让学生明确知晓数据获取的合法边界,如通过开放API、合法数据集购买等,理解个人信息与数据权益保护的重要性,从而在技术创新与法律合规的轨道上行稳致远。

五、结语

三元互动治理框架对教育数字化领域具有至关重要的启示。教育数据不仅是特殊的生产要素，更关涉受教育权、人格发展与教育公平等核心价值。教育数据的抓取与滥用，不仅扰乱市场秩序，更可能侵蚀学术诚信，甚至威胁未成年人身心健康。因此，教育领域的治理不应缺席于上述宏观讨论，反而应成为实践“三元协同”治理的先行区：立法上需加快制定教育数据分类分级标准与专属伦理规范；司法应在裁判中审慎权衡数据利用与教育公益、学生权益保护；教育行政部门更应主动履行监管职责，建立覆盖“事前备案、事中监测、事后问责”的全流程管理体系，并将数据合规纳入学校督导与教育评价。最终，通过构建清朗、安全、可信的教育数据生态，使技术真正服务于“人的全面发展”这一教育的根本目的。

参考文献：

- [1] 邓亮. 多重制度逻辑下我国教育数据开放的风险透视及治理路径 [J]. 教育学报, 2022, 18(5): 118-128.
- [2] 金玉梅, 陈航. 教育大数据挖掘的价值定位、现实限制与有效策略 [J]. 教育理论与实践, 2021, 41(19): 3-8.
- [3] 孙明阳, 刘春艳. 我国教育数据开放现状、问题与对策研究——以 15 个地方政府数据开放平台教育数据为例 [J]. 图书馆学研究, 2022(8): 30-39.
- [4] 赵磊磊, 张黎, 代蕊华. 智能时代教育数据风险治理: 实然困境与实践路径 [J]. 湖南师范大学教育科学学报, 2021, 20(6): 94-102.
- [5] 杨现民, 王娟, 李新. 加强国家智慧教育平台数据治理: 经验洞察与路径优化 [J]. 中国电化教育, 2023(9): 69-75.