

数据抓取行为的反不正当竞争法规制

——以“三重授权原则”的适用为视角

赵瑞祺, 肖湘*

广州应用科技学院法政学院、城乡文化发展研究中心, 广东 肇庆

收稿日期: 2026年3月31日; 录用日期: 2026年6月22日; 发布日期: 2026年6月30日

摘要

生成式人工智能的快速发展对传统数据抓取规制提出挑战, 起源于“微博诉脉脉案”的“三重授权原则”在司法实践中不断扩张适用, 从OpenAPI接口延伸至爬虫技术、从个人信息扩展至非个人信息、从获取环节覆盖至使用环节, 在保护平台投入的同时, 与数据流通政策、个人信息可携权及竞争自由产生张力。人工智能训练所需的海量公开数据难以实现事先授权, 使该原则面临适用困境。应回归反不正当竞争法的行为规制本质, 构建类型化规则: 根据数据是否可识别及是否衍生, 区分用户同意与平台授权的不同配置; 增设人工智能训练例外条款, 明确公开数据合理使用边界; 引入透明度义务与合理使用抗辩, 实现数据流通、权益保护与技术创新的三元平衡。

关键词

三重授权原则, 数据抓取, 反不正当竞争法, 人工智能训练

The Regulation of Data Scraping under Anti-Unfair Competition Law

—From the Perspective of the Application of the “Triple Authorization Principle”

Ruiqi Zhao, Xiang Xiao*

School of Law and Politics, Research Center for Urban and Rural Cultural Development, Guangzhou College of Applied Science and Technology, Zhaoqing Guangdong

Received: March 31, 2026; accepted: June 22, 2026; published: June 30, 2026

*通讯作者。

Abstract

The rapid development of generative AI challenges the traditional regulation of data scraping. The “Triple Authorization Principle”, originating from the Sina Weibo v. Maimai case, has been expansively applied in judicial practice—from OpenAPI to crawling technologies, from personal to non-personal information, and from data acquisition to utilization. While protecting platform investments, this expansion conflicts with data circulation policies, the right to data portability, and competitive freedom. AI training requires massive public data that can hardly obtain prior authorization, rendering the principle difficult to apply. The solution lies in returning to the behavioral regulation nature of anti-unfair competition law and establishing a typological framework based on data attributes. For identifiable primary data, only user consent is required; for non-identifiable public primary data, no platform authorization is needed; for identifiable derivative data, dual authorization is required; for non-identifiable derivative data, only platform authorization is required. Additionally, an exception clause for AI training, transparency obligations for data sources, and a fair use defense should be introduced to balance data circulation, rights protection, and technological innovation.

Keywords

Triple Authorization Principle, Data Scraping, Anti-Unfair Competition Law, AI Training

Copyright © 2026 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

2024年起, ChatGPT、DeepSeek 为代表的生成式人工智能发展迅猛,重新引起了对数据资源的重视。大模型训练对海量公共数据集的大规模采集有着极高的要求,对传统的数据采集规定产生了巨大的冲击——AI企业大量收集平台数据用于模型训练是否构成不正当竞争,是数字经济法治建设的焦点问题之一。在数据采集的问题上,“三重授权原则”具有开创性意义,2016年“微博诉脉脉案”首次提出了这一原则,即第三方通过 OpenAPI 接口获取用户信息时,需要同时获得用户许可、平台允许以及用户对平台许可的再次认可等三项授权。此后数年间,该原则适用范围不断扩大:从 OpenAPI 接口到爬虫技术,从个人信息到非个人信息,从获取数据到使用环节,这种不断扩大的使用范围虽然保护了平台投入、个人信息安全等,但也逐渐与数据流通政策、个人信息可携带权、竞争自由之间发生了矛盾,生成式人工智能的发展使这种冲突加剧,如果死板地运用“三重授权原则”进行 AI 训练就会造成“数据锁定”,阻碍技术进步和行业发展。

本文的研究方向主要集中在数据抓取行为反不正当竞争法规制方面,从“三重授权原则”出发,思考传统规则在人工智能时代所遇到的困境以及类型化方面解决问题的思路,运用规范分析、案例分析、比较法研究的方法为数据抓取行为的合法性判断提供更加灵活的视角。

2. 问题的提出:传统“三重授权原则”的形成与扩张

2.1. “三重授权原则”的司法起源

“三重授权原则”出现于 2016 年的“新浪微博诉脉脉案”中。该案的基本事实是:脉脉软件利用

OpenAPI 接口抓取新浪微博上的用户资料, 在抓取过程中仅取得了用户对微博的授权, 而未获得用户对脉脉的授权, 也没有得到新浪微博平台的授权。北京知识产权法院在二审判决中提出了“三重授权”的要求, 即第三方利用 OpenAPI 获取用户资料时, 需要遵守“用户授权”+“平台授权”+“用户授权”的三重规则。

这项规定的合理性在于以下三个方面。首先是为了保护平台的信息投入。新浪网等在用户信息的收集、加工、管理上花费了大量的精力和资金, 建立了自己的数据库并形成了有市场价值的数据库, 应当受到相应的法律保护; 其次是保护用户的个人信息权利, 用户对自己的个人信息有控制权, 第三方获取信息须经用户授权同意, 这也是符合《个人信息保护法》¹规定的“告知-同意”规则; 最后是规范市场交易秩序, “三重授权”制定了明确的规则, 减少了交易成本, 避免了“搭便车”行为对平台商业模式的破坏。

2.2. 适用范围的扩张

2.2.1. 从 OpenAPI 向爬虫模式的扩张

“三重授权”仅适用于 OpenAPI 接口调用场景——这是一种基于平台明确授权的数据获取方式。然而, 后续案件开始将该原则适用于爬虫技术, 即未经平台明确授权, 通过技术手段自动抓取网页公开数据的情形。这一扩张的典型案件包括“大众点评诉百度案”“淘宝诉美景案”等, 该案中, 百度爬取大众点评的用户评论用于百度地图, 法院虽未直接援引“三重授权”, 但实质考量了平台授权缺失与用户同意的关系, 体现了从 OpenAPI 向爬虫场景的规则迁移。法院在判断爬取行为的正当性时, 不同程度地借鉴了“三重授权”的考量因素^[1]。

2.2.2. 从个人信息向非个人信息的扩张

“三重授权”主要适用于可识别的个人信息。但随着实践的发展, 该规则被应用到用户的行为数据(如浏览历史、搜索历史), 以及数据商品(如淘宝的“生意参谋”)等非个人信息领域。这说明法院越来越重视对各大网络平台投入的保护, 但也引起了质疑——非个人信息是否需要这么严格的授权要求?

2.2.3. 从获取环节向利用环节的扩张

“三重授权”考察的是获取数据的行为本身是否正当。但是像“微信诉抖音多闪案”这样的一些案件已经把它扩展到数据的后期使用上, 也就是即便取得数据的方式合法, 如果后期的使用超过授权的范围, 则有可能构成不正当竞争。这使得“三重授权”由“入口”监管变为“全流程”监管。

2.3. 扩张适用引发的问题

2.3.1. 与数据流通政策的张力

2022 年颁布的《关于构建数据基础制度更好发挥数据要素作用的意见》(“数据二十条”)²提出了“促进数据合法合理流通利用”的观点, 主张发挥数据作为生产要素的价值最大化。但是过多强调“平台授权”也会造成“数据孤岛”——平台利用对数据的控制权来阻止竞争对手获得数据, 从而保持自己的市场份额, 这与数据流动的政策目的背道而驰。据中国信息通信研究院《数据要素发展报告(2025)》³统计, 超过 60%的企业表示“数据获取成本高”是主要障碍, 因平台过度控制数据形成的“数据孤岛”效应导致的效率损失每年估计超过 2000 亿元。

2.3.2. 与个人信息可携权的张力

《个人信息保护法》第 45 条赋予用户个人信息可携权, 即用户有权将其个人信息转移到其他平台。

¹https://www.cac.gov.cn/2021-08/20/c_1631050028355286.htm

²https://www.gov.cn/zhengce/2022-12/19/content_5732695.htm

³<http://www.caict.ac.cn/kxyj/qwfb/ztbg/202511/P020251128616212191150.pdf>

如果用户指示平台把自己的个人信息给第三方, 平台是否可以“未经平台授权”的理由拒绝? 如果严格执行“三重授权”, 那么平台的授权就会变成是用户行使可携权的一个限制, 这就与个人信息保护法的立法目的背道而驰了[2]。

2.3.3. 与竞争自由的张力

“三重授权”在赋予平台数据控制权的同时, 也赋予其排挤潜在竞争者的能力。平台可能利用技术措施(如设置 robots 协议)或法律手段, 阻止竞争者获取其公开数据, 从而削弱市场竞争。这种“数据锁定”效应在数据驱动型市场中尤为明显, 可能引发反垄断法层面的关切。

3. 传统规则的适用困境: 人工智能时代的挑战

3.1. 生成式人工智能的数据需求特征

2024 年以来, 生成式人工智能得到了快速发展。以大型语言模型(LLM)为代表的 AI 系统, 其性能的提高与训练数据集的大小以及质量密切相关。该技术特点导致 AI 企业对数据的需求具有三个特点。

第一, 大数据的海量性。大规模模型的训练通常需要 TB 或者 PB 级别的数据量。比如, GPT-3 在训练过程中使用了大约 45TB 的文本数据, 大概有几千万本图书的大小, 如此大的数据量需求使得取得平台许可或者得到用户同意在技术上无法实现。

第二, 数据来源广泛化。AI 训练数据一般来自互联网公开的网页、社交媒体、学术论文等不同的渠道, 包含数十亿个不同的数据源。如果要求 AI 企业在训练前逐个地与每个数据源进行授权商谈, 将造成不可估量的交易成本。

第三, 抓取方式的自动化与规模化。AI 公司通常使用网络爬虫工具, 自动、大批量地抓取公开网页信息, 在获得数据的方式上已经与传统的获取方式有所不同——不是为了再现或者取代原网站的功能, 而是把这些信息当作训练数据, 用以研究开发新的技术产品。

3.2. 典型争议的呈现

在中国, 自 2025 年以来也发生多起涉及 AI 训练的案件, 某 AI 平台因采集社交媒体的信息训练其模型被平台的运营者诉至法院, 控告其构成不正当竞争。虽然最终判决还没有下来, 但是矛盾焦点已经凸显: AI 公司可以在没有获得平台许可的情况下收集公开的数据吗? 用户对于自己的信息有什么权利? 实质是: 传统的“授权 - 使用”的模式是否适用于 AI 训练这一种新的方式? 当技术进步和数据权益产生冲突时, 法律应该怎样选择[3]?

3.3. “三重授权”在 AI 场景的适用困境

3.3.1. 公开数据的“平台授权”问题

AI 训练所需数据多为公开网页数据。根据传统理解, 公开数据意味着权利人放弃了排他性控制, 允许公众自由获取。然而, “三重授权”的逻辑却是: 即使是公开数据, 未经平台授权抓取仍可能构成不正当竞争。这引发了一个根本性质疑: 公开数据到底意味着什么? 平台是否有权对公开数据主张排他性控制?

3.3.2. 用户同意的实现困境

AI 训练数据中可能会存在个人信息, 但是 AI 公司很难在训练之前得到每个客户的明确授权。如果严格地遵守《个人信息保护法》中“告知 - 同意”的规定, AI 训练将会停止, 即使采用了匿名化的方法, 但是由于技术上的限制使得完全匿名化不是总能做到的。

3.3.3. 合理使用抗辩的适用争议

著作权法上对于 AI 训练合理使用的问题一直有争论。在不正当竞争法上是否也应设置相应的“合理使用”抗辩？AI 训练是否构成“转换性使用”——即对数据的利用目的(训练 AI 模型)与原网站的目的(提供信息服务)完全不同，所以不能算作不正当竞争，这个问题还有待解决。

3.4. 利益衡量的复杂性

AI 研发环境下的利益平衡，远远比传统的数据抓取争端复杂得多，至少涉及到三方利益的均衡。一是推动 AI 技术进步的公共利益。生成式 AI 被认为是新一轮科技革命和产业革命的最重要的驱动力，过于限制数据抓取行为不利于 AI 技术的进步发展，影响国家的综合实力。二是保护平台投入的私人利益。平台在进行数据收集、整理、维护等工作上花费了大量的人力物力财力，应当得到相应的回报。如果随意让 AI 公司免费获得平台的数据，可能会降低平台创新的动力。三是用户的权益不可忽视。用户对于自己的信息享有权利，这一权利不能被平台之间或者是 AI 企业之间的商业竞争所忽视。不论是平台的授予，还是 AI 企业的获取都要经过用户的同意之后再行进行。

4. 理论基础的反思与重构

4.1. 从“权利思维”到“行为规制”：反不正当竞争法的功能回归

“三重授权原则”的扩张适用，在方法论层面蕴含着一种“权利思维”，即把数据看作平台独有的财产，未经授权就是侵犯权利。但是这种思维与反不正当竞争法的制度价值取向是冲突的，反不正当竞争法的本质是行为规制法，它的主要目的是维护市场的正常竞争环境，不是为了建立专属的财产权。正如有人所说，反不正当竞争法所保护的“法益”是间接性的，其保护的如何取决于行为本身的正当性而不是基于一种假设的权利归属。

将“权利思维”引入反不正当竞争法，有两个方面的风险。一方面，使平台处于实际的数据垄断地位，可以利用反不正当竞争法排除潜在竞争对手，这与反垄断法的目的相悖。另一方面，减少了对行为本身的审查，只要是“未经授权”的数据获取就被视为违法，忽略了抓取行为的目的、方式、影响等重要因素。所以，重塑数据抓取规则的第一步就是回到反不正当竞争法的行为监管属性上来，将讨论的焦点转移到“是否经过授权”的问题上。

4.2. 数据分类的理论基础：不同数据类型的法律保护强度差异

数据并不是同质的资源，其法律性质因其产生过程、可识别性、价值构成等因素的不同而不同，在建立类型化的应用规则时，需要先建立数据分类的理论基础。第一，分为原生数据和衍生数据。原生数据是指由用户直接提供或通过其行为产生的数据，在其权利归属上应当更多地考虑用户的意见。衍生数据是指平台对原生数据进行加工、整理、分析之后得出的具有更高附加值的产品，对平台的付出应给予更多的支持。这与洛克的劳动赋权理论相契合，平台付出劳动创造了价值，就应当获得相应的法律保护。第二，分为可识别数据与不可识别数据。可识别数据涉及个人信息利益，适用《个人信息保护法》的严苛标准，不可识别数据则不涉及个人隐私，数据流通的障碍应相对较小。这也是在数据保护中遵循比例原则的体现，即保护的程度应与权利的价值成正比。第三，分为公开的数据和非公开的数据，公开的数据表明权利人愿意将其置于公共领域之中，其排他性的权利应受到限制。非公开的数据包含了权利人的商业秘密或者竞争优势，应该受到更好的保护。这也是财产法中的“公开即放弃”的原理，权利人通过公开行为，放弃了部分的排他性的控制权^[4]。

4.3. 利益平衡的新框架：流通、保护、创新的三元协调

传统“三重授权”模式背后的利益平衡逻辑是“平台保护优先”，这在数据市场发展初期是具有现实意义的。但是随着数据要素成为基本生产要素，利益平衡的方面更加多样化了，我们需要有新的分析思路来兼顾到三组基本利益。第一，数据流动与数据保护的均衡。数据是生产要素，其价值就在于流动和使用。过于重视保护平台的投入就会导致“数据孤岛”的形成从而抑制了数据本身的价值。在保证平台正当权益的基础上，要为数据流动留出合理的空间。第二，用户的权利和平台的利益之间的平衡。用户是数据的原始所有者，在对自身的数据上应有的权利得到尊重。《个人信息保护法》赋予的用户可携带权、删除权等都是对平台的数据控制权进行了限制。第三，技术创新与竞争秩序的平衡。AI等新信息技术的发展需要大量的数据共享，在倡导科技创新的同时也要警惕“搭便车”的行为带来的对平台的研发投入的打击。其中的平衡重点在于如何界定“创造性的使用”和“替代性的使用”。

4.4. 类型化适用的理论正当性

类型化适用是以上述理论为基础的必然结论。它的正当性表现在三个方面：第一，符合法律适用的比例原则——对于不同性质的数据进行不同的保护程度，防止“一刀切”的过度治理；第二，适应数据市场的多元化需求——数据应用的多样性，就要求监管手段有足够的灵活性和精准度；第三，化解多种利益之间的矛盾——运用类型化分类的方式，在不同情况下可以对流通、保护、创新等价值有不同的侧重，从而达到不断的均衡的状态。

4.5. 类型化概念的实践模糊地带与界定标准

上述“原生/衍生”“可识别/不可识别”的二分法在理论上清晰，但实践中存在模糊地带，例如用户行为日志的归类争议及匿名化数据的重识别风险。为增强可操作性，本文建议引入“数据产品”概念作为补充判断标准，即平台对原始数据进行深度加工后形成的具有独立市场价值、可独立交易的数据集合，无论是否可识别均应给予更强保护；同时建立“识别性梯度判断流程”：先判断是否包含直接标识符，再判断是否可通过间接标识符合理识别个人，最后审查平台的匿名化声明与技术隔离措施，该流程可为司法裁判提供分步检验依据。

5. 域外经验的比较法考察

5.1. 欧盟模式：《人工智能法案》的规则设计

欧盟 2024 年正式颁布《人工智能法案》⁴。该法案对于数据治理提出了两大基本原则，一是高风险 AI 系统的训练数据需要符合“数据治理”的要求，包括对数据来源的审查、对数据偏见的发现及消除等；二是重视对训练数据的透明性责任，要求 AI 服务商出具完整的技术文档，说明数据的来源、选择条件、处理方式等。

欧盟并没有对 AI 训练进行数据抓取的行为做出一般情况下的事先授权要求，而是通过透明度义务及管理制度的手段达到监管目的。这种模式符合“规范而不是禁止”的原则，既认可了 AI 训练对于数据的需求，又通过程序性规定保证了数据来源的合法性。对于我国来说，欧盟的做法也值得我们借鉴，在面对新技术出现而产生的对数据的需求时，立法者应当首先考虑如何设定合理的规则体系，而不是一味地否认数据采集的合法性。欧盟以透明度义务替代事先授权的思路值得借鉴。中国可在相关规章中引入训练数据来源公示制度，避免事先授权对 AI 创新的过度阻碍。

⁴<https://chinawto.mofcom.gov.cn/article/jsbl/zszc/202409/20240903537033.shtml>

5.2. 美国模式：司法判例的演进

美国对数据抓取的规制主要由法院在审理案件过程中作出的裁判决定。在 *LinkedIn v. HiQ Labs* 案件中，第九巡回上诉法院规定，对于公开的数据的访问，如果访问行为不违反用户协议、不破坏技术手段，则不构成违法。判决认为，平台不能以反不正当竞争为由，禁止他人抓取其公开的数据^[5]。

美国经验告诉我们合理使用规则在数据抓取领域有很大的发挥余地，中国应当在反不正当竞争法上寻找相关的“合理使用”抗辩理由。美国经验提示中国应区分反不正当竞争法、版权法、个人信息保护法的不同功能，不宜将反不正当竞争法作为数据产权的替代工具。同时，可借鉴合理使用规则的灵活性，在司法实践中探索“转换性使用”或“实质性替代”抗辩。

5.3. 日本模式：开放数据政策的制度保障

日本采取特殊的开放数据政策，通过法律规定鼓励数据流动及应用。《官民数据活用促进基本法》确立了政府数据开放的原则，《著作权法》⁵修正案则规定了 AI 训练数据合理使用规则——仅不以欣赏或盈利为目的，AI 训练就不构成著作权侵权。此政策的出发点在于体现日本“让数据流通带动创新”的基本理念。日本的经验是：数据治理不仅是一个法律问题，也是一个产业政策问题，立法者应该摒弃传统的“授权 - 侵权”的二元对立思维，以更开放的心态对待数据资源的使用。日本将数据流通上升为产业政策的思路值得借鉴。中国应在《反不正当竞争法》⁶中明确增设“AI 训练例外”条款，并区分非商业科研与商业性使用，给予不同力度的豁免。

6. 规则重构的制度路径

6.1. 类型化适用规则的确立

6.1.1. 可识别的原生数据：只需用户同意，无需平台授权

指用户的头像、昵称、职业信息等由用户直接提供的能够识别个人的信息，它的合法性在于：用户是这些信息的源头，拥有对数据的所有权；《个人信息保护法》所建立的个人可携带权保障了用户可以随意下载数据；平台仅仅是在用户授权的情况下收集这些信息，没有权力阻止用户合法地拥有这些数据。

6.1.2. 非可识别的原生数据：公开数据无需平台授权

这些数据包括用户的浏览历史、搜索历史等不能直接判断个人的行为信息，适用规则需要分为公开、非公开两种情况：如果数据是公开的，可以正常采集，防止平台利用技术手段来“圈住”公开的数据；如果数据是不公开的(即需要登录才能访问)，那么绕过平台的技术限制来获取就应视为违法。

6.1.3. 可识别的衍生数据：需用户同意与平台授权双重确认

此类数据包括用户画像、精准标签等平台加工后仍可识别个人的信息。双重授权规则的正当性在于：此类数据既承载平台的分析投入，又高度关联用户隐私，需平衡双方权益。

6.1.4. 非可识别的衍生数据：只需平台授权，无需用户同意

此类数据包括数据产品、分析报告等完全由平台投入形成且不关联个人的成果。平台对此类数据享有财产性权益，可自主决定是否授权他人使用。

6.2. “用户同意优先”的具体适用场景

个人信息转移场景下，用户根据《个人信息保护法》第 45 条要求平台将其个人信息转移到第三方时，

⁵https://www.japaneselawtranslation.go.jp/ja/laws/download/4207/05/s45Aa000480504ja14.0_r3A52.docx#26#5

⁶http://www.npc.gov.cn/npc/c1773/c1848/c21114/fbzdzjzfxd/fbzdzjzfxd002/202507/t20250702_446413.html

平台不能以“没有平台授权”为由拒绝。在此情况下,用户的指令本身就是充分的合法性依据;而对于声誉性的数据,例如:电子商务店铺评分、社交媒体影响力指数等,在用户跨平台展示的情况下,应允许第三方抓取这些信息,保证用户的信誉可以跨平台转移;对于用户生成的内容,比如:用户发布的文章、照片、视频等,如果不形成对原平台的实质性替代,则应该尊重用户的意思,让他们可以将其授权给第三方使用。

6.3. AI 训练场景的特殊规则

AI 训练中的数据使用具有很强的“转变性”,即目的(训练模型)与原网站的目的(提供资讯服务)截然不同。因此,本文提出以下特殊规定。增设“AI 训练例外”条款,在《反不正当竞争法》或有关司法解释中明确:AI 训练满足以下条件的,不构成不正当竞争——爬取对象为公开数据;爬取过程未实质性破坏平台技术保护措施;后续使用不构成对原网站核心市场的实质性替代。为进一步精细化适用,应区分模型类型与使用目的:基础大模型(foundation model)训练因数据需求海量、来源广泛,可适用例外条款;垂直领域模型(application-specific model)若涉及平台核心数据且形成直接竞争,则从严审查。非商业性科研训练可完全豁免“三重授权”,仅需履行数据来源公示义务;商业性训练则应在例外条件外增设合理补偿机制[6]。同时,明确“实质性替代”的判断指标:是否再现原平台核心内容、是否减少用户对原平台的访问频率或停留时间、是否影响原平台的核心收入来源(如广告收入、订阅用户数)、AI 产品功能是否与原平台存在直接竞争关系。若上述指标多数为否,则不宜认定为实质性替代。该例外条款的目的在于给 AI 创新提供法律上的确定性。规定数据来源的公示制度,不需要 AI 公司在训练之前获得所有数据源的同意,但要有公示其数据来源的义务。AI 公司要制定一个数据来源列表,注明其数据来源、选择标准、处理方式等等,让权利人可以行使监督和维权的权利。区别商业性和非商业性的使用,对非商业性的科研活动给予更大幅度的豁免,促进学术交流。对商业性使用设置合理的补偿机制,防止“搭便车”行为对平台激励的削弱[7]。

6.4. 司法裁判规则的优化

司法裁判不应仅以“三重授权原则”为依据,而是基于数据类型、获取方式、使用目的等作出区别对待的裁判结论。加大“实质性替代”标准的适用力度,即使未经平台授权,如果抓取行为不构成对原平台的市场替代,就应当认为是正当的行为,这有利于区分“创造性利用”和“替代性使用”。还可以运用“合理使用”抗辩,借用著作权法中的合理使用的抗辩理由,在反不正当竞争诉讼中允许被告提出合理使用抗辩,综合考虑使用目的、使用方式、对原平台的影响等因素。例如,在“大众点评诉百度案”中,法院正是基于百度地图的评价功能实质性替代了用户访问大众点评的必要性,从而认定构成不正当竞争。这体现了“实质性替代”标准对类型化规则的矫正功能。

6.5. 行业自律与技术标准的引入

明确 robots.txt 等技术规范的法律效力——遵守 robots 协议的抓取行为可推定为正当,违反者则可以推定为违法,但是也存在特殊情况的突破,比如维护公共利益、保障用户权益等等。其次,促进行业协会出台数据抓取的规定,确定抓取频率、请求间隔、数据使用范围等等标准,给市场主体一个可以预测的行为准则。

7. 结论

本文以“三重授权原则”为切入点,对数据抓取行为的反不正当竞争法规范进行全面剖析。研究发现,该原则源自“新浪微博诉脉脉案”,在司法实践中历经了由 OpenAPI 到爬虫模式、由个人信息到非个人信息、由获取环节到使用过程三个方面的扩展,在保护平台投入的同时,和数据流通、个人信息可

携带性以及竞争自由产生了矛盾冲突。生成式人工智能的出现使得这一矛盾更加突出。解决之道在于回归反不正当竞争法对行为规范的本质要求, 在数据属性的基础上形成类型化规则: 原始可识别数据只需要用户同意即可, 原始公开的不可识别数据则无需平台授权, 衍生的可识别数据则需要双方授权, 非可识别的衍生数据则只需要平台授权即可, 加入 AI 训练的豁免条款, 增设合理使用的抗辩理由, 制定出一套符合时代要求的规则框架。本文的不足之处在于数据分类的界限、技术的可行性以及域外经验的本土化问题有待加强。以后的数据确权立法、AI 的发展以及国际规则的融合都将为数据治理提供一个新的制度框架。

参考文献

- [1] 黄武双, 邱思宇. 数据流通视角下公开数据抓取的反不正当竞争法规制之反思[J]. 竞争政策研究, 2025(3): 38-52.
- [2] 黄武双, 邱思宇. 行为区分视角下公开数据抓取的反不正当竞争法规制[J]. 科技与法律(中英文), 2025(3): 1-13+25.
- [3] 沈贵明, 刘源. 数据抓取行为反不正当竞争法规制困境与对策[J]. 中国流通经济, 2021, 35(1): 89-96.
- [4] 盛茹. 数据抓取行为的反不正当竞争法规制研究[J]. 中国物价, 2023(8): 86-89.
- [5] 梅芷涵. 数据抓取行为的反不正当竞争法规制研究——以脉脉案为切入点[J]. 经营与管理, 2022(5): 24-29.
- [6] 林婧, 陈琳. 数据抓取行为的反不正当竞争法规制[J]. 电子科技大学学报(社会科学版), 2021, 23(6): 28-35.
- [7] 陈沛. 数据流通与利用中的“三重授权”原则——再评大数据引发不正当竞争第一案[J]. 上海市经济管理干部学院学报, 2020, 18(1): 57-64.