我国网络爬虫技术的数据合规法律风险及其完善建议

——以我国网络爬虫511个案例的实证分析为出发点

严嘉欢　钱书豪　王宇婷　巩　浩*

内容摘要：网络爬虫作为数据时代重要的数据分析、数据挖掘工具，获得了互联网企业的青睐，然而网络爬虫的肆意生长还暗含刑事违法的风险。最高人民检察院发布的《涉案企业合规典型案例（第三批）》其中重点提到了网络爬虫刑事合规案件，足以看出我国对于网络爬虫数据合规的重视性。通过对于我国2013—2022年的511个网络爬虫案件进行数据实证分析，发现目前我国企业使用网络爬虫技术数据合规的现存三个方面问题，即事前网络爬虫爬取行为的授权性未知、事中网络爬虫爬取数据范围的模糊性、事后网络爬虫爬取数据使用的不可控。针对上述问题提出我国企业使用网络爬虫数据合规的完善建议，包括事前强调网络爬虫爬取数据的授权前提、事中明确爬虫爬取数据范围的合法性、事后确保爬取数据使用的合规性。

关键词：网络爬虫　数据合规　个人信息　数据安全　实证分析　企业合规

一、我国网络爬虫数据合规案例的法理分析

（一）爬虫与ROBOTS协议——善意与恶意爬虫

爬虫善意与恶意区分的标准便是ROBOTS协议。ROBOTS协议在我国被认定为商业道德，它既非防火墙，又无强制执行力，是否能够遵守该协议产生善意与恶意爬虫之分。善意爬虫能够遵守ROBOTS协议，最常见的便是各大搜索引擎。而恶意爬虫则会肆无忌惮地抓取数据，常造成网站崩溃，同时也常作为一些企业爬取数据实施不正当竞争的工具，侵犯个人信息权、财产权等重要权益。从现实来看，爬虫抓取数据并侵害他人权益行为屡次出现，网络爬虫已沦为违法犯罪工具，成为个人信息泄露、不正当竞争行为的罪魁祸首，而现有ROBOTS协议也并不足以阻止恶意爬虫行为，因而对爬虫的规制成为企业合规的重点。

（二）爬虫与数据竞争的厘清

互联网与平台经济时代的商业竞争，企业之间所面临的最大竞争便是数据竞争。为了在数据竞争

*严嘉欢，华东政法大学国际金融法律学院硕士研究生；钱书豪，西北政法大学公安学院本科生；王宇婷，西北政法大学民商法学院本科生；巩浩，西北政法大学民商法学院本科生。

严嘉欢负责总体框架与撰写，钱书豪负责校对修改与撰写，王宇婷负责数据分析与撰写，巩浩负责收集资料与撰写。

中争得上风，企业将爬虫技术充分介入，对数据展开激烈争夺，但同时也引发了垄断之嫌。数据竞争纠纷，主要集中于在先获取数据的企业与对数据再使用的企业之间。前者为获取数据付出了一定的成本，因而不会轻易将劳动成果交予他人使用。但数据掌握在少数企业手中势必会造成垄断的僵化局面，无法合理配置市场各要素。后者为了获取数据，需要经过重重授权，这也无疑增加了企业的成本。在这样的矛盾下，后者会转而以爬虫技术爬取大量数据，但这却对前者构成威胁。

网络经济具有外部性，[1]消费者选择产品的过程中，容易被锁定在某一个产品上，这种外部性也容易成为企业扩大生产规模的着眼点。在数据竞争中，这一特性发挥出了最大的作用，企业会为了获取这些相关数据，使用爬虫技术，成则可以垄断市场，败则面临生存危机。网络爬虫技术能高效地抽取有价值信息，[2]而数据蕴含的巨大价值诱发了一系列侵犯数据的新型行为，其中以网络爬虫最为典型。[3]当下，类似的数据爬取纠纷在我国主要是通过反不正当竞争法第2条[4]来解决的，是否构成不正当竞争的判断标准是诚实信用原则或商业道德。但该标准具有很大的不确定性，实践中转而以利益平衡机制相衡量。在该机制下，利益平衡是一个动态博弈的过程，各方对数据的掌握程度主导着利益边界的调整。对何种数据附加以何种获取和使用方式，既能保护数据来源主体的权益，从根本上保障数据获取的延续，又能最大限度地保护在先获取数据企业的利益，同时又不妨碍对数据再使用的企业对数据的获取，充分实现数据共享与流通，是必须要考虑的问题。

（三）企业合规与爬虫之间的现实冲突

2022年5月，我国首例短视频平台网络爬虫案宣判，[5]标志着法律对爬虫技术领域的规制不断深入，同时也意味着爬虫工具的爬取行为的合法性需要质问与反思。爬取行为本身作为一种中性技术手段，其合法与否的定性与其所爬取的范围、对象高度相关。以短视频平台网络爬虫第一案为例，案中爬虫服务器通过入侵服务器的方式对平台中用户的用户名、UID、签名及评论等个人信息进行搜集汇总，从而形成精准用户画像。丁某被判处刑罚的关键在于其所要求的网络爬虫在入侵他人未公开服务器的情景下实行爬取行为，并对服务器中未公开用户信息进行搜集爬取。由此观之，爬取行为的技术使用和法律合规、保障公民数据权益之间的持续性平衡。

（四）企业网络爬虫合规问题成因

第一，爬虫作为一种自动化的工具能够有效提高数据价值的转化，其技术行为本身为中性，但是其作为双刃剑，一方面从事实上能够促进数据的再次利用，另一方面在利用的程度评价上存在合法与否的价值标准。

第二，在规制爬虫案件中，适用的传统法律为刑法、民法典侵权责任编、反不正当竞争法，其中作为补充的为近年新出台的个人信息保护法、数据安全法。现有个人信息保护法与数据安全法仅构建出数据使用规范的基本框架，而并未对具体爬虫爬取行为有直接规制。现对爬虫的爬取行为进行规制的紧密相连的仅有网站所有者使用的Robots.txt文件（《互联网搜索引擎服务自律公约》第7条）。[6]由此，在爬虫使用过程中便存在难以对违规侵犯个人信息行为、不正当竞争行为进行有效规制。具体而言有

〔1〕网络外部性可以从不同的角度来理解，主流的观点倾向于消费者层面来认识：当一种产品对用户的价值随着采用相同产品或可兼容产品的用户增加而增大时，即出现网络外部性。网络外部性意味着原有用户得到了产品中所蕴含的新增价值而无须为这一部分价值提供相应的补偿，因而是属于经济学中所阐述的正的外部性的一个特例。

〔2〕参见杨松令、刘梦伟、张秋月：《中国金融科技发展对资本市场信息效率的影响研究》，《数量经济技术经济研究》2021年第8期。

〔3〕参见童云峰：《大数据时代网络爬虫行为刑法规制限度研究》，《大连理工大学学报（社会科学版）》2022年第2期。

〔4〕反不正当竞争法第2条：经营者在生产经营活动中，应当遵循自愿、平等、公平、诚信的原则，遵守法律和商业道德。本法所称的不正当竞争行为，是指经营者在生产经营活动中，违反本法规定，扰乱市场竞争秩序，损害其他经营者或者消费者的合法权益的行为。本法所称的经营者，是指从事商品生产、经营或者提供服务（以下所称商品包括服务）的自然人、法人和非法人组织。

〔5〕参见《全国首例短视频平台领域网络“爬虫”案宣判》，载https://mp.weixin.qq.com/s/XUdWskKHoMKKGno7BE4lgQ，2022年5月11日。

〔6〕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。

如下几个问题：

首先，对网络爬虫合法类型界定不清。网络爬虫作为一个中性的技术工具，其性质随着使用者的善意与恶意也在变化。数据安全法第23条〔7〕虽然对恶意爬虫有禁止性规定，但是并未对爬虫爬取行为是否为恶意进行有效判别。公开网页行为是否属于合法正当方式、窃取的边界在何处，这些问题均未有效提及。刑法设置侵入计算机信息系统罪、侵犯公民个人信息罪固然违法，而刑罚与正当行为之间的民事责任、行政责任的边界并未厘清，对网络爬虫的善意与恶意并无明确的区分责任。反不正当竞争法作为解决爬虫相关问题的重要行政责任规制依据，在解决与互联网相关的纠纷中，仅对强制插入、跳转等传统问题进行处理，并未对爬虫、算法方面的数据竞争问题提供有效规制，在遇到此类问题时，仅有反不当竞争法第2条作为原则性条款可以适用，而面对井喷式的新兴互联网算法纠纷显得力不从心。

其次，对个人敏感信息的界定不清。爬虫爬取过程中，其爬取对象是任意的，仅有目标网站中存在的ROBOTS文件作为软性规制，而爬虫可以随时冲破ROBOTS协议，对其所需要的信息进行爬取。并且网站中ROBOTS协议属于网站管理者所设定的社区公约，与个人信息保护法中个人信息的规制范围并不一定重合，由此爬虫爬取个人信息范围存在诸多交叉灰色地带。同时，个人信息保护法第4条仅提及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。该条并未明确说明个人信息的具体含义与范围，同时在不同情境下个人信息的含义并不相同。不同网站对个人信息程度要求不同，如证券业务网站需要精确评估客户资产年龄甚至投资能力，而一般的娱乐网站仅需在客户同意的情形下收集其娱乐偏好，在这样不同场景中，信息的敏感与否存在较大差异。而在爬虫爬取之后个人信息的范围与保护方式更加难以定义。例如用户在某社交网站中自愿公开自己的个人信息，而并非意味着同意其个人信息在他处爬取后再利用。虽然个人信息保护法第28条〔8〕对爬虫爬取客户敏感信息行为有更加细致的规制，但爬虫爬取行为认定存在较大障碍，导致法律责任落实不足，追责措施滞后。由此，客户个人信息边界难以捋清，爬虫爬取对象的范围、侵权范围难以规制。

最后，对用户知情同意边界界定不清。根据个人信息保护法第30条，〔9〕在用户签署个人信息服务中，其收集信息条款多为对此应用APP或网页中信息授权，而并非赋予爬虫爬取的权利，在爬虫爬取之后个人信息并未脱敏，未达到合理使用的标准。而在互联网公开平台中，一旦用户信息公开即存在被抓取的危险，对于此类危险仅有ROBOTS协议作为微薄的权利保护屏障，明显存在知情同意后被滥用的灰色危险地带。

二、我国网络爬虫数据合规纠纷类型案件实证数据分析

笔者从裁判文书网汇总了全国2013—2022的10年间关于网络爬虫数据合规纠纷类型案件的裁判文书，共计511份，其中包括了民事案件425件，刑事案件72件，行政案件12件，管辖案件4件。笔者从案号、案由、案件类型、审理程序、审理法院、案件发生区位、被诉对象、原告诉讼请求、被告抗辩理由、法律依据和判决结果11个方面对每个案件进行了主要信息的筛选，最后汇总为一个类案信息检索记

〔7〕数据安全法第23条：任何组织、个人收集数据，应当采取合法、正当的方式，不得窃取或者以其他非法方式获取数据。

〔8〕个人信息保护法第28条：敏感个人信息是一旦泄露或者非法使用，容易导致自然人的人格尊严受到侵害或者人身、财产安全受到危害的个人信息，包括生物识别、宗教信仰、特定身份、医疗健康、金融账户、行踪轨迹等信息，以及不满十四周岁未成年人的个人信息。只有在具有特定的目的和充分的必要性，并采取严格保护措施的情形下，个人信息处理者方可处理敏感个人信息。

〔9〕个人信息保护法第30条：个人信息处理者处理敏感个人信息的，除本法第十七条第一款规定的事项外，还应当向个人告知处理敏感个人信息的必要性以及对个人权益的影响；依照本法规定可以不向个人告知的除外。

录表。

网络爬虫是以自动运行编程为主的一种计算机程序，其特点是可以自动、长久地向服务器请求数据。本文综合考虑原告、被告、法院三方在案件中持有的态度，针对511份判决书进行归纳分析，从案由、案件类型、审理程序、审理法院、案件发生区位、被诉对象、原告诉讼请求、被告抗辩理由、法律依据和判决结果十个方面，客观公正地对企业、公民违规爬取数据进行评析。目的是通过一系列可视化数据，来剖析当前公民与企业，企业自身间在网络安全方面存在的矛盾，并根据现有法规精确分析网络爬虫涉及的法律责任，以此来预防违规网络爬虫，并为网络爬虫营造宽松良好的环境。需要指出的是，由于涉及数据抓取案件的数量较为有限，因此，实证分析得出的结果可能存在一定的片面性，但是结论对于表述当前国内数据抓取案件审判的大致倾向具有一定的代表性，对于后续司法实践和理论研究也具有一定的借鉴和参考作用。

下面将511份有效案件通过SPSS统计软件和Excel软件进行可视化分析。

（一）基础系数模型的建立与分析

选取案由、案件类型和判决结果三个因素作为案例的基础系数，并进行细致客观地分析。

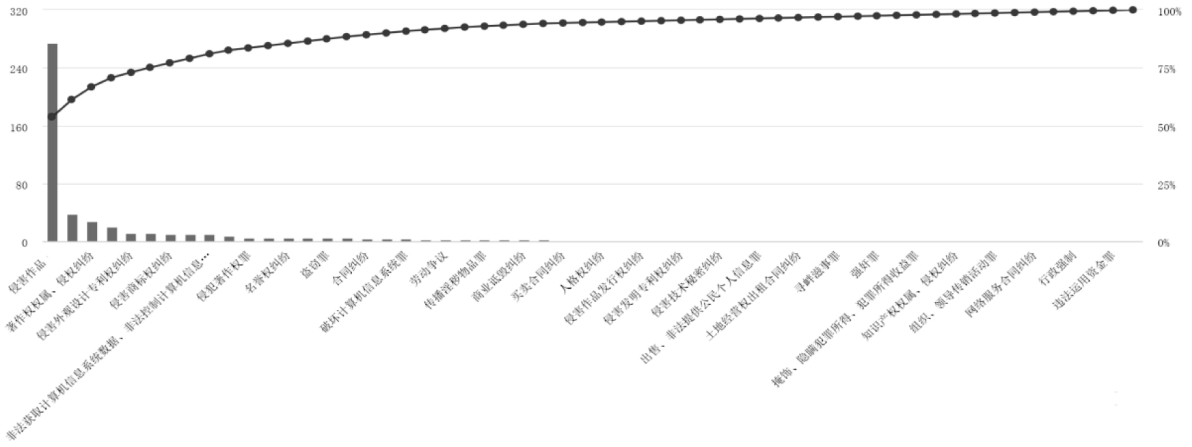

图1　案由系数帕累托统计图

针对案由，如图1，结合“二八原则”进行帕累托图分析可知：侵害作品信息网络传播权纠纷、不正当竞争纠纷、著作权权属、侵权纠纷、侵犯公民个人信息罪、侵害外观设计专利权纠纷、商业贿赂不正当竞争纠纷、侵害商标权纠纷、计算机软件开发合同纠纷共8项为“至关重要项”，该8项累计占比为79.06%。表明了虽然爬虫犯罪原因多样，但是当前爬虫犯罪接近80%来源于这八项纠纷。而非法爬虫活动的刑事法律责任主要集中于非法获取计算机信息系统数据罪与侵犯公民个人信息罪，理论上亦可涉及非法侵入计算机信息系统罪、非法控制计算机信息系统罪以及提供侵入、非法控制计算机信息系统程序、工具罪等。虽然，企业之间的数据抓取既能满足用户的检索需求，是推动行业进步的网络技术手段，实现互利共赢，也会是爬取他人数据资源、抢夺市场优势地位、侵犯其他经营者权益的竞争工具。

除此之外，非法获取计算机信息系统数据、非法控制计算机信息系统罪、诈骗罪、侵犯著作权罪、制作、复制、出版、贩卖、传播淫秽物品牟利罪、名誉权纠纷、服务合同纠纷、盗窃罪、行政裁决、合同纠纷、提供侵入、非法控制计算机信息系统程序、工具罪、破坏计算机信息系统罪、其他行政行为、劳动争议、劳动合同纠纷、传播淫秽物品罪、侵害作品署名权纠纷、商业诋毁纠纷、技术服务合同纠纷、买卖合同纠纷、产品销售者责任纠纷、人格权纠纷、侵害企业名称（商号）权纠纷、侵害作品发行权纠纷、侵害作品广播权纠纷、侵害发明专利权纠纷、侵害商业秘密纠纷、侵害技术秘密纠纷、其他（质量监督）、出售、非法提供公民个人信息罪、名称权纠纷、土地经营权出租合同纠纷、定作合同纠纷、寻衅滋事罪、开

设赌场罪、强奸罪、技术委托开发合同纠纷、掩饰、隐瞒犯罪所得、犯罪所得收益罪、擅自使用他人有一定影响的企业名称、社会组织名称、姓名纠纷、知识产权权属、侵权纠纷、确认不侵害著作权纠纷、网络侵权责任纠纷、组织、领导传销活动罪、网络服务合同纠纷、职务侵占罪、行政强制、计算机软件著作权转让合同纠纷、违法运用资金罪、非法侵入计算机信息系统罪等共48项为“微不足道项”，该48项累计占比为20.94%。

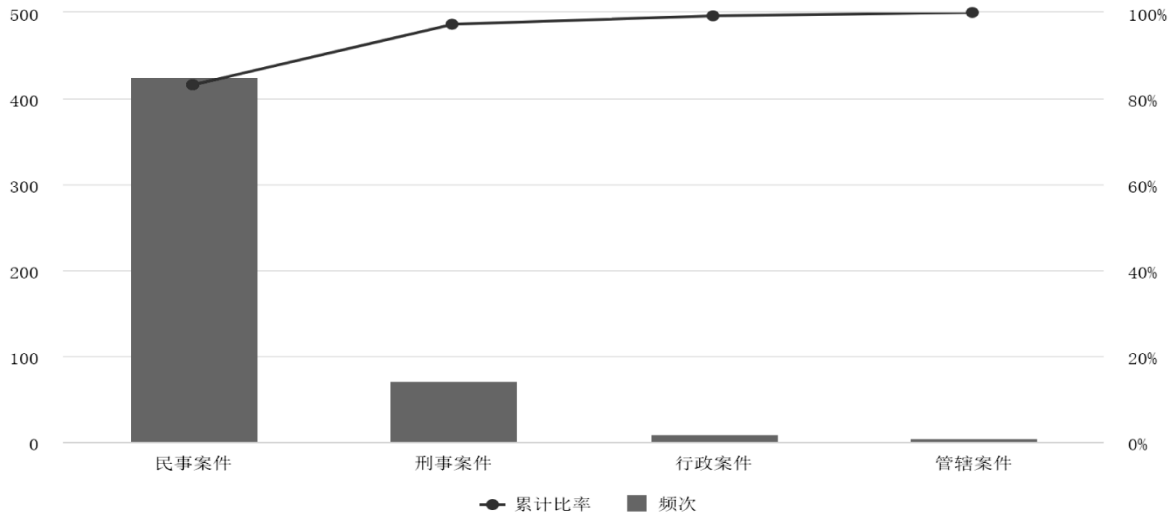

图2 案件类型系数帕累托统计图

针对案件类型，如图2结合“二八原则”进行帕累托图分析可知：民事案件共1项为“至关重要项”，该1项累计占比为83.17%。除此之外，刑事案件，行政案件，管辖案件等共3项为“微不足道项”，该3项累计占比为16.83%。由此可见，网络爬虫被用作侵犯他人合法权益的工具对社会产生了负面影响。首次，从网络平台搭建方权益的角度来看，网络爬虫作为爬取信息的工具将会导致网络平台运行过载，甚至将会导致被爬取网站直接崩溃而其他用户无法正常访问的问题，这将严重影响平台的正常经营运行与流量盈利。其次，从平台用户本身的个人信息权益的角度来看，网络爬虫作为爬取工具将会导致用户的个人信息提取与泄露，这将会严重侵犯个人隐私。当爬虫爬取对象是政府网站的情形下，非法侵入内网获取政府工作秘密。[10]在用户本身是企业的情况下，还会导致企业的商业秘密的泄漏。

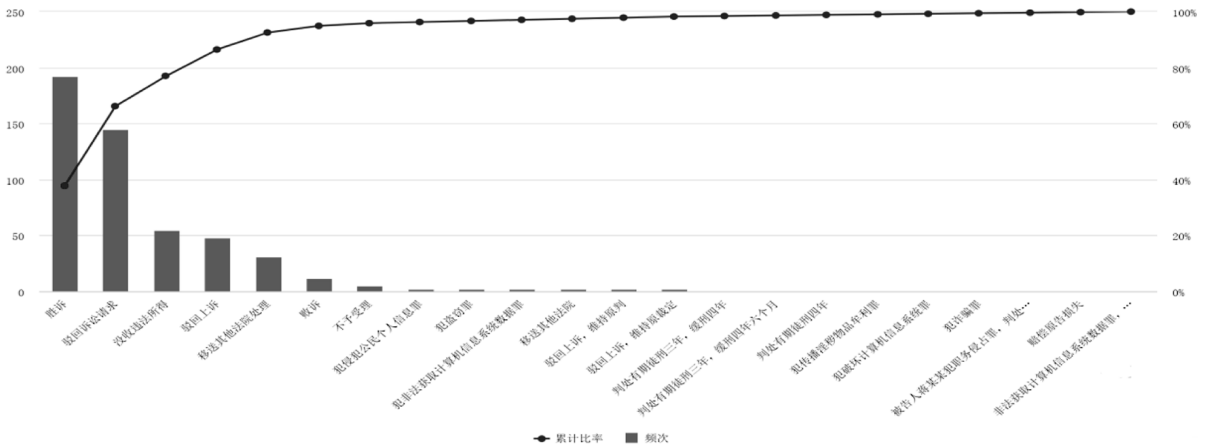

图3 胜败诉系数帕累托统计图

〔10〕参见苏宇：《网络爬虫的行政法规制》，《政法论坛》2021年第6期。

而对于判决结果，结合“二八原则”，如图3分析可知：胜诉，驳回诉讼请求，没收违法所得共3项为“至关重要项”，该3项累计占比为77.01%。除此之外，驳回上诉，移送其他法院处理，败诉，不予受理，犯盗窃罪，犯非法获取计算机信息系统数据罪，移送其他法院，驳回上诉，维持原判，驳回上诉，维持原裁定，犯侵犯公民个人信息罪，判处有期徒刑三年，缓刑四年，犯破坏计算机信息系统罪，犯诈骗罪，判处有期徒刑三年，缓刑四年六个月，被告人蒋某某犯职务侵占罪，判处有期徒刑六个月，被告人退赔的赃款予以发还，判处有期徒刑四年，赔偿原告损失，非法获取计算机信息系统数据罪，判处有期徒刑三年，缓刑三年，犯传播淫秽物品牟利罪等共19项为“微不足道项”，该19项累计占比为22.99%。

由此可见，法院在受理案件之后，因为程序问题或者实体问题，又大量地做出了驳回起诉或者驳回部分诉求的判决，综上可知，我国关于网络安全和个人信息保护方向的法律亟待完善，而公民对于个人信息保护的意识依旧有所欠缺。

（二）重要相关系数的解读

笔者从七个方面中，摘取“案由”“被诉对象”和“被告抗辩理由”这三个作为因变量，并与“案由类型”“行政区域”“原告诉由”“法院观点”“适用法条”“胜败诉”，这七个自变量分别进行相关性分析，在SPSS软件中采用Pearson相关性分析，取样本数N=406，在0.01水平（双侧）及0.05水平（双侧）检测是否显著相关，并以此来分析大型平台对不同类型案件中是否具有相关性，判断爬虫数据合规机制中是否具有集中纰漏性等内容。

表1　Pearson系数相关表

	案由	被诉对象	被告抗辩理由
案件发生区位	0.020	0.077	0.039
原告诉讼请求	0.108*	0.075	–0.263**
审理程序	–0.193**	–0.205**	–0.054
案件类型	–0.217**	–0.250**	0.121**
审理法院	0.147**	0.066	0.036
法律依据	0.300**	0.188**	0.079
判决结果	–0.254**	–0.162**	–0.211**
*p<0.05　**p<0.01			

从表1可知，利用相关分析去研究案由，被诉对象，被告抗辩理由分别和案件发生区位，原告诉讼请求，审理程序，案件类型，审理法院，法律依据，判决结果共7项之间的相关关系，使用Pearson相关系数去表示相关关系的强弱情况。具体分析可知，案由和原告诉讼请求之间的相关系数值为0.108，并且呈现出0.05水平的显著性，案由和审理法院之间的相关系数值为0.147，并且呈现出0.01水平的显著性，案由和法律依据之间的相关系数值为0.300，并且呈现出0.01水平的显著性，因而说明案由与原告诉讼请求、审理法院和法律依据之间呈现出显著性正相关数值。除此之外，案由与案件发生区位之间的相关关系数值并不会呈现出显著性（p>0.05），意味着案由与案件发生区位之间并没有相关关系。而被诉对象与审理程序、案件类型和判决结果之间呈现出显著性负相关数值。具体来看，被诉对象和审理程序之间的相关系数值为–0.205，并且呈现出0.01水平的显著性，因而说明被诉对象和审理程序之间有着显著的负相关关系。被诉对象和案件类型之间的相关系数值为–0.250，并且呈现出0.01水平的显著性，被诉对象和判决结果之间的相关系数值为–0.162，并且呈现出0.01水平的显著性，此外，被诉对象与案件发生区位，原告诉讼请求，审理法院共3项之间的相关关系数值并不会呈现出显著性（p>

0.05),意味着被诉对象与案件发生区位,原告诉讼请求,审理法院之间并没有相关关系。

由此可见大数据背景下个人信息采集路径的多元化、大数据信息平台的开放共享性及数据加密措施的可破解性,使得数据交易与数据共享亦会面对隐私保护与信息安全的挑战。〔11〕

(三)回归模型的分层深化

分层回归用于研究自变量(X)增加时带来的模型变化,通常用于模型稳定性检验,中介作用或者调节作用研究。笔者通过变量与变量内部因素之间进行交叉回归分析,以了解彼此之间具体影响因素以及是否会有影响。

表2 分层回归模型简化表

	分层1	分层2	分层3	分层4	分层5	分层6
常数	16.947** (49.451)	4.291** (3.958)	4.612** (4.194)	3.269** (3.200)	5.140** (4.658)	6.809** (5.829)
案由	−0.071** (−4.604)	−0.041** (−3.051)	−0.032* (−2.260)	−0.048** (−3.741)	−0.047** (−3.763)	−0.034** (−2.633)
审理程序		−0.062 (−0.151)	0.310 (0.714)	0.280 (0.701)	−0.017 (−0.044)	−0.066 (−0.172)
案件类型		6.345** (13.900)	6.224** (13.511)	5.569** (12.964)	6.043** (13.953)	5.773** (13.361)
审理法院			−0.022 (−1.772)	−0.025* (−2.143)	−0.020 (−1.773)	−0.022* (−1.989)
案件发生区位			0.061 (0.770)	0.090 (1.242)	0.078 (1.115)	0.093 (1.357)
原告诉讼请求				0.113** (9.218)	0.090** (7.189)	0.087** (7.061)
被告抗辩理由					−0.096** (−6.216)	−0.090** (−5.834)
被诉对象					−0.002 (−0.803)	−0.001 (−0.399)
法律依据						−0.026** (−3.897)
样本量	466	466	466	466	466	466
*R*2	0.044	0.328	0.337	0.440	0.486	0.502
调整*R*2	0.042	0.324	0.330	0.433	0.477	0.492
*F*值	*F* (1,464)= 21.199,*p*=0.000	*F* (3,462)= 75.180,*p*=0.000	*F* (5,460)= 46.741,*p*=0.000	*F* (6,459)= 60.223,*p*=0.000	*F* (8,457)= 53.914,*p*=0.000	*F* (9,456)= 51.098,p=0.000
△*R*2	0.044	0.284	0.009	0.104	0.045	0.017
△*F*值	*F* (1,464)= 21.199,*p*=0.000	*F* (2,462)= 97.750,p=0.000	*F* (2,460)= 3.072,*p*=0.047	*F* (1,459)= 84.971,*p*=0.000	*F* (2,457)= 20.016,*p*=0.000	*F* (1,456)= 15.185,*p*=0.000
因变量:判决结果						
*p<0.05 **p<0.01 括号里面为t值						

从表2可知,本次分层回归分析共涉及6个模型。模型1中的自变量为案由,模型2在模型1的基础上加入审理程序,案件类型,模型3在模型2的基础上加入审理法院,案件发生区位,模型4在模型3的基础

〔11〕参见徐宗新、陈沛文:《数据红利与信息危机——兼论网络爬虫的罪与罚》,《上海法学研究》2021年第1卷。

上加入原告诉讼请求，模型5在模型4的基础上加入被告抗辩理由，被诉对象，模型6在模型5的基础上加入法律依据，模型的因变量为：判决结果。

将案由作为自变量，而将判决结果作为因变量进行线性回归分析，从上表可以看出，模型R方值为0.044，意味着案由可以解释判决结果的4.4%变化原因。对模型进行F检验时发现模型通过F检验（F=21.199，p<0.05），也即说明案由一定会对判决结果产生影响关系，以及模型公式为：判决结果=16.947-0.071*案由。

案由的回归系数值为-0.071，并且呈现出显著性（t=-4.604，p=0.000<0.01），意味着案由会对判决结果产生显著的负向影响关系。因而案由均会对判决结果产生显著的负向影响关系。

针对模型2，其在模型1的基础上加入审理程序，案件类型后，F值变化呈现出显著性（p<0.05），意味着审理程序，案件类型加入后对模型具有解释意义。另外，R方值由0.044上升到0.328，意味着审理程序，案件类型可对判决结果产生28.4%的解释力度。具体来看，审理程序的回归系数值为-0.062，但是并没有呈现出显著性，意味着审理程序并不会对判决结果产生影响关系。案件类型的回归系数值为6.345，并且呈现出显著性（t=13.900，p=0.000<0.01），意味着案件类型会对判决结果产生显著的正向影响关系。

针对模型3，其在模型2的基础上加入审理法院，案件发生区位后，F值变化呈现出显著性（p<0.05），意味着审理法院，案件发生区位加入后对模型具有解释意义。另外，R方值由0.328上升到0.337，意味着审理法院，案件发生区位可对判决结果产生0.9%的解释力度。具体来看，审理法院的回归系数值为-0.022，但是并没有呈现出显著性，意味着审理法院并不会对判决结果产生影响关系。案件发生区位的回归系数值为0.061，但是并没有呈现出显著性，意味着案件发生区位并不会对判决结果产生影响关系。

针对模型4：其在模型3的基础上加入原告诉讼请求后，F值变化呈现出显著性（p<0.05），意味着原告诉讼请求加入后对模型具有解释意义。另外，R方值由0.337上升到0.440，意味着原告诉讼请求可对判决结果产生10.4%的解释力度。具体来看，原告诉讼请求的回归系数值为0.113，并且呈现出显著性（t=9.218，p=0.000<0.01），意味着原告诉讼请求会对判决结果产生显著的正向影响关系。

针对模型5：其在模型4的基础上加入被告抗辩理由，被诉对象后，F值变化呈现出显著性（p<0.05），意味着被告抗辩理由，被诉对象加入后对模型具有解释意义。另外，R方值由0.440上升到0.486，意味着被告抗辩理由，被诉对象可对判决结果产生4.5%的解释力度。具体来看，被告抗辩理由的回归系数值为-0.096，并且呈现出显著性（t=-6.216，p=0.000<0.01），意味着被告抗辩理由会对判决结果产生显著的负向影响关系。被诉对象的回归系数值为-0.002，但是并没有呈现出显著性，意味着被诉对象并不会对判决结果产生影响关系。

针对模型6：其在模型5的基础上加入法律依据后，F值变化呈现出显著性（p<0.05），意味着法律依据加入后对模型具有解释意义。另外，R方值由0.486上升到0.502，意味着法律依据可对判决结果产生1.7%的解释力度。具体来看，法律依据的回归系数值为-0.026，并且呈现出显著性（t=-3.897，p=0.000<0.01），意味着法律依据会对判决结果产生显著的负向影响关系。

在当下的时代，技术日新月异，欧洲实施了如此严苛的数据保护条例尚且还被质疑会阻碍科技的创新，“只要市场存在，即使监管再严格，也总会有人会因为利益去铤而走险”，滥用“网络爬虫”技术，一旦达到一定的程度，就可能触犯刑法，就传统法益领域，可能侵害公民个人信息、知识产权等，就新型法益领域，网络数据系统安全等就可能成为“网络爬虫”技术侵害的法益。〔12〕

然而网络爬虫作为普遍运用的一种信息搜集技术，本身是技术中立的，如果该技术被合法、正当地使用，则有助于打破数据壁垒，促进数据资源的流通和共享，可有力推动社会经济的发展。如果该技术被滥用，会侵害他人利益，损害他人技术创新的积极性，并给使用者带来一系列的法律风险。若企业

〔12〕参见陈军标、杨兰：《“网络爬虫”技术的法律规制》，《上海法学研究》2020年第12卷。

遭受相应的刑事处罚、行政处罚或承担民事责任，企业除了高昂的经济损失，还会使商誉受损，不利于企业的持续经营。随着数据资源对于社会主体的重要性日益增强，引发的问题也越来越多，对于立法、司法的挑战也在不断革新。立足于当前的中国国情，如何对网络爬虫技术进行规制，是许多部门法需要共同面对的难题。〔13〕

三、我国企业使用网络爬虫技术的现存问题分析

通过对判决书进行Pearson相关性分析，根据所得结果可知，在大数据背景下，个人信息采集路径多元，且大数据信息平台具有开放共享性，其数据加密措施又可破解，因而个人信息被泄露与盗用的风险极高，在数据交易与数据共享的过程中，隐私保护与信息安全皆会面临挑战。而爬虫所衍生的问题，贯穿于爬虫技术使用前后的全过程。

（一）事前问题：网络爬虫爬取行为的授权性未知

在遵守 Robots 协议的前提下，判断爬虫获取数据行为合法性边界可以参考以下两点：

1.爬取的数据为公开数据还是非开放数据

就公开数据而言，第三方在抓取和使用过程中在“最少、必要”的合理范围内，无须得到经营该用户信息平台的授权，反之，则需要得到授权。就开放数据而言，第三方抓取网站的开放数据，不仅需要平台的授权许可，也需要信息提供方即用户的许可，适用“三重授权”的模式。如果目标网站有反爬取协议，应严格遵守网站设置的 ROBOTS 协议。可以说，无论从保护网民隐私还是尊重版权内容的角度，遵守 ROBOTS 协议都应该是正规互联网公司的默之举，任何违反 ROBOTS 协议的行为都应该为此付出代价。

2.爬取行为是否损于被爬取方

数据爬取方采用的爬取方式是否实质上妨碍被爬取方的正常经营（如是否干扰网站的正常运营，是否破坏系统正常运行，是否导致服务器崩溃损坏等），是否不合理增加被爬取方的运营成本。爬取行为不应妨碍网站的正常运行。企业应当合理控制爬取的频率，尽可能避免过于频繁地抓取数据，特别是如果超过了《数据安全管理办法（征求意见稿）》明确规定的“自动化访问收集流量超过网站日均流量三分之一”的要求，就应当严格遵守网站的要求，及时停止数据抓取。

（二）事中问题：网络爬虫爬取数据范围的模糊性

数据爬取范围的问题一共有两个方面。第一个方面，企业在使用的网络爬虫技术爬取的数据是否属于公开获取的数据判断方面仍存模糊性问题。在网络爬虫案中，公开的这些信息数据其实是被公众所能够获取并且所知悉的，将这些案件认定为以数据信息保密性为主线的侵犯商业秘密罪抑或非法获取计算机信息系统罪，都可能偏离其不法行为的本质。抓取公开数据，原则上不应予以入罪。〔14〕且根据前述数据分析结果来看，案由也会对判决结果产生一定的影响，若将爬虫所引发的纠纷全部归入刑事案件，则会面临败诉的结果。再者，在我国的司法实践中，普遍将使用网络爬虫获取公开信息这一公开性质的特点作为被告企业的抗辩事由。然而伴随目前爬虫技术与反爬虫技术的不断对抗与升级，公开数据的认定也将会变得更为复杂。例如反爬虫验证措施之下所能获取到的数据是否属于公开数据，仍然需要予以明确规定。目前反爬虫验证措施已经有了较大的升级与进步，现如今较为权威的反爬虫措施以谷歌为代表，谷歌所采用的选中图片中某一特定类型物品的验证措施能够有效地阻止大量网络爬虫的入侵。但这也会带来一个问题，即在正常情况下，平台用户可以顺利通过反爬虫测试，获

〔13〕参见陈军标、杨兰：《“网络爬虫”技术的法律规制》，《上海法学研究》2020年第12卷。

〔14〕参见吴卫：《明确越界网络爬虫行为的刑事处罚边界》，载https://www.spp.gov.cn/spp/llyj/202202/t20220215_544538.shtml，2022年11月28日。

取到平台内部存储的特定的数据。那么这种越过反爬虫测试所获得的向用户公开的数据是否属于公开数据，仍然存在一定的疑问。如果今后有一款网络爬虫能够顺利越过此类反爬虫算法测试，获取到这些公开的信息数据，那这些使用网络算法的企业是否可以用公开数据这一条款进行免责抗辩，仍然需要加以明确规定。

第二个方面，该企业使用的网络爬虫技术爬取的数据是否属于特定权限访问的特殊数据仍然存在一定的模糊性，例如是否属于个人敏感信息等范畴。以网络爬虫的刑法规制为例，网络爬虫本为获取搜索信息的正常工具，但由于犯罪分子对爬虫技术进行“非法变种”，导致利用网络爬虫的行为涉嫌构成侵犯公民个人信息罪及其他犯罪。〔15〕在一般情况下，检察机关判断该网络爬虫所爬取数据是否属于个人敏感信息更多是从该网络爬虫爬取到数据这一瞬间来决定的，但是该网络爬虫功能是否会对于加密信息进行去匿名化、去标识化的信息复原功能，往往难以判断。

（三）事后问题：网络爬虫爬取数据使用的不可控

数据爬虫技术是支撑数字经济的一种手段，但对其利用的目的可能不再单纯，对于爬取数据的使用也并不一定符合企业数据合规与数据经济良善发展的目标。无论是善意爬虫还是恶意爬虫，对数据的使用都存在不可控的数据垄断风险以及数据泄露风险，〔16〕兼具技术性与复杂性，而这也会涉及诸多纠纷，根据判决书数据分析结果，爬虫所引发的案件类型以民事和刑事为主，且民事案件占据绝大部分比例。

信息时代，数据被大大赋值，一些企业对所收集的数据采取保护主义，更有甚者，企业会利用爬虫技术恶意爬取对其有利的信息并据为己有，形成庞大的数据资源库，企图垄断市场。在互联网经济外部性的作用下，企业通过爬虫技术掌握消费者偏好，将目标群体一一吸引至自身的市场中，以扩大市场份额，最终实现独霸的局面。这便是数据垄断。

数据垄断的后果是企业对数据的随意处置。具体而言，企业通过对数据进行垄断，为自身构建起数据寡头的霸权地位，精准掌控相关信息，吸引大量客户群，迅速在市场上立足，并攫取其他市场的消费者，使得其他企业难以存续而被兼并，由此形成数据垄断企业独霸市场的终局形态。同时，为了进一步增强竞争力，这些企业之间也会进行一些附带不正当竞争的目的的数据交换，其后果不可控，个人信息来源主体的重要权益极易受到侵犯。〔17〕利益驱使下，垄断企业内、外部之间利用数据这一具有重要价值的载体实施关联犯罪，非法买卖数据，泄露信息，形成犯罪链条。根据判决书数据分析，当前爬虫犯罪所涉案由便主要有不正当竞争纠纷、著作权权属、侵权纠纷、侵犯公民个人信息罪、商业贿赂不正当竞争纠纷等纠纷，而非法爬虫活动的刑事法律责任主要集中于非法获取计算机信息系统数据罪与侵犯公民个人信息罪。〔18〕

ROBOTS协议所设定的商业道德只不过是一种效力极弱的软约束，故爬虫在爬取数据时并不存在真正意义上的障碍。企业利用恶意爬虫在获取数据时便已不具备正当目的，因此更无法期待其能够依法使用数据。而利用善意爬虫的企业虽然在最初遵循了ROBOTS协议，但这并不意味着其也能完全按照法律和道德使用数据。一旦企业将所获取的数据进行非法出售，其他企业便能通过不正当的交易获取这些数据并进行二次传播，导致数据危机爆发，形成不可逆转的蝴蝶效应。另外，一些企业在主观上或许并无泄露数据的目的，但是其系统存在漏洞，数据库遭到攻击，也会造成数据被动泄露的后果。对于企业而言，数据泄露导致商业秘密等关乎企业生存的重要信息处于暴露状态，同时也会让客户丧失信心，导致企业失去竞争优势，生存面临困境。对于作为数据来源者的个人而言，数据泄露直接导致其个人信息和隐私泄露，生活安宁权受到破坏。对于国家和社会而言，数据泄露还会形成犯罪链条，加

〔15〕参见林雨佳：《刑法司法解释应对新型科技犯罪的逻辑、立场与路径》，《东方法学》2022年第3期。

〔16〕参见王鹏飞：《数据安全导向下企业刑事合规保护体系建构》，《天津师范大学学报（社会科学版）》2022年第6期。

〔17〕参见丁晓东：《论数据垄断：大数据视野下反垄断的法理思考》，《东方法学》2021年第3期。

〔18〕参见苏宇：《网络爬虫的行政法规制》，《政法论坛》2021年第6期。

剧社会动荡，危害国家安全。

从判决书数据分析结果可以看出，公民对于个人信息保护的意识依旧有所欠缺，且就现有法律和治理措施来看，也并不能对企业使用数据形成很好的威慑，而爬取数据以及使用数据本身存在技术性操作，这也导致很难预测企业该如何使用数据并对其加以控制。对获取的信息加以传播、利用或改造，轻则侵害民事权益，重则有可能涉及诸多犯罪。但由于民法等前置法的惩罚效果微弱，刑法又存在谦抑性，故只能在网络爬虫行为产生严重社会危害而无刑罚以外手段进行规制的情形下起到惩治效果，[19]这也就会使得企业常行走在违法犯罪的边缘，在不被法律所严惩的范围内任意使用爬取的数据，而这实质上已经造成了巨大的危机。判决书数据分析结果表明，法院在受理因爬虫引发的案件之后，又因为程序或者实体问题，大量作出了驳回起诉或者驳回部分诉求的裁判。由此可见，并非所有受到网络爬虫侵害的纠纷均能得到有效地解决，网络安全、个人信息保护和爬虫数据使用之间仍然未能形成有效的平衡局面，如何规制“网络爬虫”技术的使用，仍旧是亟待解决的难题。

四、我国企业使用网络爬虫技术的合规建议

（一）事前合规：强调网络爬虫爬取数据的授权前提

爬虫采用的技术是否具有侵入性地突破数据访问控制，法律上是否突破网站或APP的ROBOTS协议限制。需要合理限制抓取的内容。

第一，在设置抓取策略时，应注意编码禁止抓取视频、音乐等可能构成作品的、明确的著作权作品数据，或者针对某些特定网站批量抓取其中的用户生成内容；在使用、传播抓取到的信息时，应审查所抓取的内容，如发现属于用户的个人信息、隐私或者他人的商业秘密的，应及时停止并删除。20对于内部系统数据，严格禁止侵入。

第二，可以创新数据处理的技术手段，如利用隐私计算技术处理原始化数据，通过“可用不可见”的方式使用原始数据，在保障原始数据安全的情况下，充分发挥其价值。[21]又如，可以利用区块链技术的去中心化，对数据进行加密处理，增强数据不易篡改的特性，在事前进行处理，有效控制数据在使用过程中可能发生的风险。[22]

（二）事中合规：明确爬虫爬取数据范围的合法性

在刑事合规方面，还需要在事前确保企业所采用的网络爬虫算法所爬取的数据范围的合规性质。然而，实际上刑法刑事罪名的规制与其他部门法的著作权侵权、个人信息侵权、不正当竞争行为等其他行为侵权规制仍然存在较大的差异性。在爬虫爬取数据范围的行为过程中，要明确区分爬取行为的合法性认定、处罚的程度性认定。由此，达到保护合法数据行为、促进我国网络生态繁荣发展；对不同情节的违法行为进行公正处罚，做到违法必究的同时也注重罚当其罪。因此建议从事中的行为角度的违法性评价角度保障网络爬虫算法爬取数据范围的合规性。具体从技术角度分析可以分为以下两个方面：

第一，明确网络爬虫非法爬取数据的处罚范围，以确保互联网企业在合规的范围之内爬取相关数据。以爬取个人信息方面的网络爬虫为例，相关企业应当充分考虑该网络爬虫在数据收集过程中是否遵循了知情同意以及最小必要数据收集的原则。企业还需要对爬虫技术的运用进行合规监控，在直接爬取个人信息数据时要履行告知同意义务。[23]由于网络爬虫功能复杂，目前我国在刑法领域对于网

〔19〕参见阮林赟：《网络爬虫刑事违法的立场、标准和限制》，《河北法学》2021年第7期。

〔20〕参见刘鹏：《利用网络爬虫技术获取他人数据行为的法律性质分析》，《信息安全研究》2019年第6期。

〔21〕参见唐林垚：《数据合规科技的风险规制及法理构建》，《东方法学》2022年第1期。

〔22〕参见刘艳红：《智慧法院场景下个人信息合规处理的规则研究》，《法学论坛》2022年第6期。

〔23〕参见孙跃：《数字经济时代企业数据合规及其构建》，《湖北社会科学》2022年第8期。

络爬虫的规制更多是从数据获取的角度进行的，例如侵犯公民个人信息罪、非法获取计算机信息系统数据罪等，而不是设定专门的网络爬虫过度爬取数据罪这一类工具特定化罪名。这也就导致了诸多罪名都可以规制网络爬虫这一工具，企业使用网络爬虫进行数据爬取的过程中容易造成诸多罪名之间的想象竞合。因此，为了进一步推动网络爬虫刑事合规的实现，我国刑法应当在事前加强考虑对于网络爬虫非法爬取数据范围的限定，而不是一味追求将诸多刑法罪名施加于该互联网公司，形成全方位的多重刑法罪名的事后规制。在明确非法爬取数据范围方面，首先我国刑法应当从网络爬虫技术的不断发展与创新的角度出发，重新审视在一定反爬虫限制措施例如谷歌图片类型选择等措施之下，大量获取公开开放数据是否能够纳入刑法的规制范围之内。其次，应当充分考虑限制获取的数据的限定范围，尤其是在今后网络爬虫不断进步的过程中，刑法也应当针对网络爬虫去匿名化、去标识化等加密信息反向破解领域的新型犯罪进行相应的规制。

第二，适当考虑将平台ROBOTS协议、自律公约等行业内部自律协定纳入刑事数据合规考察范围之内，以改善平台ROBOTS协议、自律公约在刑事合规中视若无睹的尴尬局面现状。根据个人信息保护法、民法典等法律法规的规定，权利人是否同意是判断行为人获取个人信息行为非法性的重要标准，而具体到网络爬虫领域，网页信息权利人的意愿体现在ROBOTS协议上。〔24〕目前，在刑法规制领域没有形成一个非常明确的网络爬虫可爬取数据范围的相关规定。在网络爬虫爬取数据范围合规性方面，除去法律法规的监管规制之外，平台也使用ROBOTS协议一类的自律协议与公约等文件形式对于网络爬虫的数据排除范围进行了规制。在最高人民检察院发布的网络爬虫合规典型案例中，Z公司违规使用网络爬虫违反E平台商户端协议，通过爬虫程序大量获取E公司存储的订单信息等数据。〔25〕在此案中，就存在Z公司的平台商户端协议这一行业内部自律协议。最典型的技术规范即目前互联网行业惯常采用的ROBOTS协议，是一种用于规制网络爬虫获取网站信息的文本文件。〔26〕实质上平台所制定ROBOTS协议也具备一定科学性，从经济价值而言，让平台来制定相关的ROBOTS数据采集范围的协议是最为科学、最节省经济资源的，因为不同平台中包含的不同数据具备较大的差异性。我国检察机关也应当充分考虑ROBOTS在刑事合规方面的作用。在ROBOTS协议适用方面，美国已有相关判例。美国法院认定某一网络爬虫使用行为为非法行为的典型案例是2000年的 Ebay v. Bidder´s edge案件。在该案中，原被告双方就围绕数据权属及网络爬虫展开了讨论。〔27〕在本案中，Bidder´s Edge在未经许可的前提下使用爬虫爬取并复制了Ebay网页的内容。法官认为，在本案中Bidder´s Edge使用网络爬虫的行为因并未遵守Ebay制定的ROBOTS协议，因而Bidder´s Edge所使用的网络爬虫行为超出了E-bay网站的授权权限范围。〔28〕最终，美国法院认为被告通过网络爬虫获取EBay商品数据的行为构成侵权。〔29〕分析美国在网络爬虫领域的这一典型性案例可以知道，尽管该案例并不是刑事案件，但仍然能够看出ROBOTS协议在网络爬虫爬取数据范围方面授权的决定性作用。然而事实上，目前平台所制定的ROBOTS协议在我国刑法认定过程中仍然存在较为尴尬的局面，即我国刑法并未明确ROBOTS协议在网络爬虫获取数据范围刑事定性领域中的作用，因此在这一层面，建议我国检察机关加强对于网络平台所制定的ROBOTS协议以及行业制定的自律公约等文件在数据获取范围刑事合规领域的认定作用。

（三）事后合规：确保爬取数据使用的合规性

就企业对爬取数据使用的现状而言，确保爬取数据使用的合规性是企业数据合规的一大重点。

〔24〕参见黄陈辰：《论公开信息的刑法保护》，《大连理工大学学报（社会科学版）》2022年第3期。

〔25〕参见陈瑞华：《合规顾问在有效合规整改中的作用》，《浙江工商大学学报》2022年第6期。

〔26〕参见何敏、马诗雅：《互联网企业数据不正当竞争一般条款适用逻辑之辨》，《科技与法律（中英文）》2022年第2期。

〔27〕EBAY,INC.v. BIDDER'S EDGE, INC, 100 F.Supp.2d 1058,2000.

〔28〕EBAY,INC.v. BIDDER'S EDGE, INC, 100 F.Supp.2d 1058,2000.

〔29〕参见谢宜璋：《可商品化数据的进一步厘清：概念、保护诉求及具体路径》，《知识产权》2021年第8期。

事后合规主要在于风险管控，合规治理的重点也应集中于如何预防和减少数据泄露等风险的发生与扩大。爬虫是一种技术化手段，因此对数据使用的合规管理也应当注重强化技术赋能，在满足数据不被泄露的前提下合理公开相关信息，严格对待可能被二次利用的数据，尽量减少数据在使用的过程中带来的次生损害，对各种可能存在的风险与损害进行合理预测，制定综合的应对方案。

第一，可以利用数据脱敏技术对数据进行匿名化处理，从源头上降低数据泄露带来的风险。在此意义上，将信息与主体之间割裂开来，也即进行脱敏处理，切断它们之间的联系，减少数据泄露的风险。要进行脱敏化处理，最首要的便是要确定合理脱敏范围。范围过大则会过当地限制数据价值的发挥，反之又无法起到相应的合规治理效果。在确定范围的时候，要避免“一刀切”，根据不同的情况对不同数据进行不同处理，实现数据保护与利用的动态平衡。〔30〕此外，对数据进行脱敏处理还要明确匿名化的程度，把控好比例问题，避免因过度匿名化导致数据失去价值。

第二，如果不能避免数据违规的发生，那么便要将其所产生的不良影响控制在最小限度内。为此，可以对企业使用数据制定不同的应对方案，及时减少风险的扩大。应对方案在于及时性〔31〕和全面性。数据安全危机爆发的速度和传播覆盖的范围会因互联网的催化大大增加，一旦不及时控制，就会产生无法挽回的后果。另外，从获取数据到使用数据，是技术运作的过程，而技术本身就具有复杂性和风险性，因此必须充分考虑所有可能出现的情况，制定全面的应对措施。

第三，进行实时监管，及时评估合规治理的效果，并不断对方案和措施进行完善。对于企业数据使用是否得当的监管标准，在于企业是否从形式上为数据加强了保密外观，实质上是否兼顾了数据保护和数据价值发挥的平衡。以此为指引，对数据使用情况的监管可以采取审慎监管以及内外双线监管多管齐下的方式，以数据使用的风险管控与化解为目标，客观评价风险状况，及时进行风险预测、预警以及控制，并在企业内部建立监管部门，及时反馈合规治理的情况，同时在外部设立独立的数据使用监管机构，防止企业内部因利益保护而异化导致监管失效。

结　语

网络爬虫本身是中立的技术工具，是背后使用人的意识与目的决定了爬虫到底是属于益虫害虫。从法经济学角度来说，网络爬虫主宰下的世界无论是危机四伏的“数字丛林”还是寸步难行的“数字沼泽”都不利于数字经济的发展。如果网络爬虫泛滥生长而不加以任何刑法方面的规制与限制，那么数字经济发展的世界将会成为数字丛林。消费者与平台用户在使用网络平台的时候，会发现自己的信息数据被随意获取。在这样的状态下，消费者与平台用户仿佛深陷数字丛林，而暗藏的数字爬虫就宛如丛林中的毒虫，随意侵扰、撕咬消费者。而如果网络爬虫技术被禁止使用，数字经济的发展状况又会像数字沼泽一般难以前进，没有任何人能够从中顺利通行。数据技术也难以获取相关的经济利益，这也对于数据生产要素转化为生产力本身造成了巨大的经济阻碍，不利于我国数据产业赋能政策的实现。

不可否认的是，网络爬虫作为数据挖掘的数字技术检索工具其本身能够使得庞大繁杂的数据本身转变为极具经济价值的具有归类特征的数据群。同时网络爬虫技术工具能够充分实现文本分析、数据挖掘、数据匹配、机器学习等多项理论，具有较高的经济价值。刑事规制不能一味强调摒弃、排斥网络爬虫技术工具的使用，而应当以积极的、开放的心态来面对网络爬虫工具本身。刑法的规制与以爬虫技术工具为代表的数字经济的发展之间应当存在良性平衡的发展关系，不能一味地对于爬虫技术苛以严格的刑法规制。在这样的发展状态下，刑事合规在数字经济领域方面的运用是两者平衡最好的

〔30〕参见刘艳红：《智慧法院场景下个人信息合规处理的规则研究》，《法学论坛》2022年第6期。

〔31〕参见王鹏飞：《数据安全导向下企业刑事合规保护体系建构》，《天津师范大学学报（社会科学版）》2022年第6期。

实践与调和处理方式。最高人民检察院发布的第三批企业合规典型案例中就明确包括了网络爬虫案例，这也为日后检察机关处理数字经济领域的刑事案件作出很好的指引性规定。未来我国数字经济领域刑事合规的发展方向必将是基于该企业使用的网络爬虫等数据分析处理技术，充分考虑该计算机技术获取相关数据并进行分析的算法原理的安全性，来明确是否对相关案件的被告人采取刑事措施，根据该数据分析技术的社会危害性来决定是否需要采取刑事合规措施。