人工智能法治专题

文章编号:1008-4355(2025)05-0013-17

生成式人工智能训练数据的来源治理

谷 川

(北京市水务综合执法总队,北京 100036)

摘 要:训练数据是生成式人工智能模型构建与优化的关键要素之一,也是模型内容生成品质与可信性的前提与保障。面对生成式人工智能模型开发对数据规模性与多元化的利用需求,有必要解决训练数据在不同来源场景下可能面临的法律风险。未来应当通过协调、改进或优化个人数据权益、作品数据版权以及企业数据权益等既有制度安排,最小化数据权益保护与创新利用的负外部性影响,以促进在模型开发与权益保护之间形成新的利益平衡。在保护数据既有权益的同时,法律亦应公平拓展数据资源在模型构建与优化中的潜在利用空间,为模型开发者获取来源数据提供必要的可行路径。

关键词:生成式人工智能;训练数据;法律治理;法律经济学

中图分类号:D923 **文献标志码**:A

DOI:10.3969/j.issn.1008-4355.2025.05.02 **开放科学(资源服务)标识码(OSID)**:

一、问题的提出

生成式人工智能是指具有文本、图像、音频或视频等内容生成能力的相关技术;它以数据、算法和算力为核心要素,业已成为现阶段人工智能领域最具突破性和影响力的创新技术之一。[①]相对于算法、算力,训练数据通常被认为是生成式人工智能更为基础和关键的生产要素。大规模的训练数据并非现成可得,生成式人工智能的模型开发者(以下简称模型开发者)需要通过互联网等诸多渠道加以收集、获取。鉴于数据资源呈现出的利益叠加性与主体多元化等特点[②],模型开发者在收集、

收稿日期:2025-02-25

基金项目:国家社科基金青年项目"人工智能与《民法典》双重背景下个人信息保护研究"(20CFX041)

作者简介:谷川(1982—),男,北京人,北京市水务综合执法总队工作人员,法学博士。

① 参见许雪晨、田侃、李文军:《新一代人工智能技术(AIGC):发展演进、产业机遇及前景展望》,载《产业经济评论》2023年第4期,第6-15页。

② 参见高富平:《论数据持有者权 构建数据流通利用秩序的新范式》,载《中外法学》2023年第2期,第313-314页。

获取用于模型训练来源数据的过程中①,往往会面对诸多利益相关者的数据权益“丛林”,并可能侵犯这些在先权益,进而使来源数据处于非法利用或不当利用的状态,以至于影响后续来源数据开发与利用的合法性与正当性。例如,用于模型训练的个人数据或作品数据在规模化的获取中,事先可能未得到用户或作者的同意,且在数量上也远超过现有制度的容忍限度;利用网络爬虫等自动化收集手段,大规模任意(含避开或破解技术措施等方式)爬取企业持有的数据,给持有企业带来数据安全或商业利益上的负面影响等。因此,在生成式人工智能领域,如何解决数据权益冲突,如何进一步有效激励模型开发者对数据资源的创新利用,并为其提供来源数据收集、获取方面的制度安排,便成为当下乃至未来,促进数字技术可持续发展所需回应的一个重大治理议题。

二、训练数据来源的法律风险与治理挑战

在经验层面上,训练数据来源的法律风险与治理挑战,主要是指模型开发者在收集、获取用于生成式人工智能模型构建与优化的数据过程中,对他人在先数据权益造成侵害或产生不利影响的风险,以及对既有法律制度带来的治理挑战。结合现有实践中来源数据的类型,本文将训练数据来源的场景风险与治理挑战主要限定在对个人数据权益②、作品数据版权以及企业数据权益潜在侵害等三个方面。

(一)对个人数据权益侵害的风险与治理挑战

无论是欧盟的《一般数据保护条例》(GDPR),还是我国的《中华人民共和国个人信息保护法》,均对个人数据采取了较为严格的保护模式,即明确赋予个人数据主体相应的保障性权益以及个人数据处理者的义务负担。③ 就前者而言,可将个人数据权益的保护聚焦在知情同意、自主决策以及维护保障等三个方面④;就后者而言,处理者在收集与使用个人数据的过程中,原则上除事先向个人告知并取得同意外,还应当遵循此类数据处理的“目的必要性”限制——在明确、合理的处理目的约束下,在最小范围内收集与使用数据⑤,以尽可能减少对个人数据权益的负外部性影响。在大数据与人工智能时代,特别是基于生成式人工智能的发展格局下,模型开发者在生物制药、医疗健康、智能驾驶以及金融服务等专业领域对个人数据的需求与日俱增,既有个人数据权益保护的“告知同意”规则以及“目的必要性”原则的约束,很大程度上会使模型开发者就此类数据的收集与利用面临高昂的交易成本与较大的法律风险。这不仅加大了上述制度的实施成本,更在一定程度上限缩了数据创新利用的潜在合理空间。

---

① 本文所谓的“来源数据”,是指用于生成式人工智能模型训练的各类数据,既包括直接用于模型输入的训练数据集,也涵括需经清洗、整理或标注等预处理后方可用于模型训练使用的原料数据。

② 鉴于数据是信息的一种电子记录形式或表现载体,故若无其他说明,本文对数据与信息不另行区分,相应的个人数据权益亦等同于个人信息权益。

③ 例如,《中华人民共和国个人信息保护法》第5章“个人信息处理者的义务”的规定;欧盟《一般数据保护条例》(GDPR)第4章第1节(控制者与处理者)基本义务的规定。

④ 参见谷川:《数据要素的权利界定与制度保障:基于效率的法律激励》,载《西南政法大学学报》2023年第5期,第110页。

⑤ 例如,《中华人民共和国个人信息保护法》第6条的规定;欧盟《一般数据保护条例》(GDPR)第5条第1款(b)(c)项规定。

对于“告知同意”规则,个人数据处理者逐个事先告知并取得数据主体的同意,在规模不大、有限数据量的情形下,或许具有一定的可行性(主要在于交易成本不高);但面对生成式人工智能模型训练所需的庞大数据量,无论是模型开发者收集个人原始数据,还是从其他持有者处继受取得个人数据,除面临较大的交易成本负担与法律风险外,还会对模型的训练与开发带来更多的不利影响。以继受取得个人数据的场景为例,数据持有者逐个向众多数据主体取得同意会进一步提高数据的交易成本。① 为了降低运营成本与法律风险,数据持有者自然会降低对外提供数据的意愿,并提高数据提供的对价;而对于作为需求方的模型开发者来说,鉴于其处在数据流通的下游位置,上游数据供应不足,就可能影响到模型开发者模型训练的潜在效能。

“目的必要性”原则通常强调应当以一个明确、合理的处理目的约束个人数据处理者,在能够实现处理目的的最小范围内收集个人数据,并采取对个人权益影响最小的方式使用个人数据。但在实践中,鉴于生成式人工智能模型开发的多元化与训练数据的规模性获取等技术特点,模型开发者对于个人数据的收集与使用遵循上述原则亦面临较大困境。尤其是在收集、获取用于模型训练的个人数据时,可能会因模型训练所需数据量异常庞大,导致个人数据处理的“最小范围收集”“最小影响使用”等约束机制流于形式,在实践中难以落实。

总的来看,无论是“告知同意”规则,还是“目的必要性”原则,基于传统治理理念的个人数据保护机制设计,其基本目标仍是倾向于扭转数据主体在数据处理中的劣势地位,并试图增强数据主体对其数据的控制或决策能力。但这种制度安排,又会与大数据时代的数据开发,特别是与生成式人工智能模型训练对海量数据资源的需求格格不入,尤其是在特定专业领域的模型训练中,个人数据的规模性、代表性、完整性等要求恰恰成为模型训练所必需。

(二)对作品版权侵害的风险与治理挑战

就生成式人工智能而言,高质量的内容生成离不开高质量的训练数据,而高质量的训练数据则有赖于高质量来源数据的有效供给。在不少情形下,具有较高质量的来源数据通常表现为知识产权的客体,故模型开发者在对训练数据的收集与使用过程中,往往也会涉及知识产权问题。② 随着网络与大数据处理技术的日趋成熟,越来越多的作品通过网络数据化的形式得以广泛传播,在方便作品信息传递与公众获取的同时,某种程度上也会引发更多关于网络环境中涉及作品数据收集与使用的版权纠纷问题。

在实践中,往往会出现模型开发者通过爬取方式获得作品,但未取得著作权人许可并支付相关使用费用,进而涉嫌侵犯著作权的情形。例如,研发运营生成式人工智能模型产品 ChatGPT 的美国 OpenAI 公司,在生成式预训练模型的开发与优化过程中,由于其大量爬取与利用互联网媒体平台发布的新闻和评论报道等作品,被纽约时报等多家媒体平台要求承担侵权责任并支付使用费用。③ 相

① 以《中华人民共和国个人信息保护法》第 23 条为例,数据提供方(个人数据处理者)向其他处理者提供个人信息的,应当取得个人的单独同意。

② 参见焦和平:《人工智能创作中数据获取与利用的著作权风险及化解路径》,载《当代法学》2022 年第 4 期,第 128-131 页。

③ See Thomas C. Carey, *The New York Times v. OpenAI: The Biggest IP Case Ever*, Sunstein LLP, https://www.sunsteinlaw.com/publications/the-new-york-times-v-openai-the-biggest-ip-case-ever, visited on 2025-2-1.

关诉求的正当性看似很清晰,按照传统版权治理规则,使用他人作品的,原则上需经著作权人同意,并支付相关版权使用费用。[①] 尽管诸多作品在网络媒体平台上是以公开方式呈现的,但这并不意味着模型开发者可不经作者许可而任意使用作品。对于生成式人工智能的模型开发来说,当前作品的版权保护及其合理使用规则,很大程度上难以满足模型开发者对于此类数据的利用需求,相互间产生的利益冲突愈发明显。

一方面,“授权使用”的版权交易模式会导致模型开发者数据获取的效率损失。鉴于模型开发者对数据资源的海量需求,逐一要求其与作品权利人协商并取得授权的成本较为高昂,这不仅涉及到与著作权人进行联系、沟通的信息成本,更会涉及作品的议价成本,耗费大量时间与精力。另一方面,在使用方式上,模型开发者亦有别于传统版权法意义上的作品使用。就后者而言,一般集中体现在“对作品中的独创性表达的直接或间接再现”[②],而在生成式人工智能的模型开发与优化中,若仅从技术的视角审视,对于被用于机器学习的作品,开发者只是将其作为分析与计算自然语言的“碎片化”原料,以期提高发现、掌握相关内容生成规律的概率[③],似乎难以与表达、欣赏等传统利用方式相关联。

另外,传统版权保护方案中的合理使用规则,尚不能满足生成式人工智能模型开发训练的需要。一般而言,作为版权保护限制与例外的合理使用规则,需要符合“三步检测法”[④]的要求。一是受制于“特殊事项”[⑤]的约束。从目前来看,用于模型开发的数据训练活动尚不属于合理使用法定范围中的“特殊事项”,且模型开发者对训练数据的挖掘分析活动是以海量的数据资源供给作为前置性条件的,这本身就与版权法中作品合理使用在数量、规模上的约束相悖。二是不应与作品的正常使用相冲突。就“正常使用”的边界而言,若将作品用于模型训练、挖掘分析的活动纳入“正常使用”的范围,则模型开发者对作品的使用或有碍于著作权人对其作品的正当利用。三是不得不合理地影响或妨害著作权人对作品的合法利益。鉴于模型开发者在数据预处理过程中,可能使原有的作品数据按照一定的标准或方法进行重新归类或整理等,导致原有作品内容上的变动,进而影响到作品的完整性,涉嫌对著作权人保持作品完整性权利的侵犯。

(三)对企业数据权益侵害的风险与治理挑战

无论是承载个人信息的数据,还是涉及作品内容或非作品内容(如对商家产品或服务的点评信息等)的数据,往往都由数据企业持有。这类数据通过不同商业模式的运作,形成企业的经营资源与商业(竞争)利益。这类数据往往被称为企业数据,即由市场主体在生产运营中合法生成、收集且

① 除《中华人民共和国著作权法》外,《生成式人工智能服务管理暂行办法》第7条亦明确规定生成式人工智能服务提供者(模型开发者或运营者)进行训练数据处理活动时,不得侵害他人依法享有的知识产权。

② 陶乾:《基础模型训练的著作权问题:理论澄清与规则适用》,载《政法论坛》2024年第5期,第160页。

③ 参见刘晓春:《生成式人工智能数据训练中的“非作品性使用”及其合法性证成》,载《法学论坛》2024年第3期,第70页。

④ 通常认为,“三步检测法”源于《伯尔尼公约》等国际规定,后被我国版权制度所借鉴,如《著作权法》第24条的规定。

⑤ 主要体现在对作品内容的个人欣赏、研究、学习,或新闻报道、履行公务等特定事项上的利用,且通常限定在少许、有限的使用数量与范围内。

控制的各类数据资源。[①] 从相关政策、司法判例等实践来看,对数据持有者的权益保护与企业数据的授权使用是一种基本框架。例如《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》(第2部分第5项)、《关于促进企业数据资源开发利用的意见》(第2部分)确立的企业数据权益及其授权使用框架机制。由此可见,企业数据权益及其授权使用机制的出台,在一定程度上,构成了对第三方使用者收集、使用此类数据的一种约束。

数据使用者针对企业进行数据收集的法律风险点,主要涉及的是以非合意方式对此类数据的获取行为。[②] 对于企业已设置访问权限的可(半)公开数据、不公开数据[③],传统治理思路通常是从数据安全以及不正当竞争等规制视角,对非法访问与获取此类数据的行为持否定态度。[④] 经验事实表明,上述行为不仅会对数据持有企业的数据安全带来较大风险,而且有可能对数据持有企业的商业模式等竞争利益带来不当影响,故既有治理方案中的"授权使用"机制,无论在业界还是学界,尚无太大争议。

模型开发者收集与使用企业数据的主要问题集中于企业的公开数据方面。模型开发者能否在不经数据持有企业许可的前提下获取其公开数据,以及在收集、获取的限度等问题上,相关规则仍不明确,可能导致模型开发者对此类数据处理的"两难境地"。一方面,若收集此类数据,可能存在非法或未经授权获取数据之嫌。例如,"hiQ实验室诉领英数据爬取纠纷案"(hiQ Labs, Inc. v. LinkedIn Corp.)中,原告hiQ实验室收集被告领英持有的公开用户数据,是否需要经持有企业授权这一关键问题,虽经美国联邦地区法院、上诉法院甚至最高法院等多层级的审理,目前仍没有定论。[⑤] 无独有偶,在国内司法实践中,对于企业公开数据抓取的相关纠纷,主要是在反不正当竞争法的框架下,对此类行为的正当性进行评价,而非直接明确公开数据原则上是否可被他人任意收集、获取。[⑥] 另一方面,若不收集此类数据,又难以满足生成式人工智能开发过程中对数据的规模性需求,同时也必然影响数据资源利用的效率。

总的来看,既有制度与司法实践倾向于保护数据持有者,而相对疏于对公开数据获取与使用的制度化合理指引。质言之,数据持有企业与作为第三方的模型开发者利益冲突的深层原因,主要在于企业公开数据的权益界定仍然有待完善。数据参与者之间的权益边界不够清晰,致使企业公开

---

① 一般意义上,此类数据既包括企业自行收集或生成的数据(原始取得数据),也包括从其他数据持有者处转移获取的数据(继受取得数据)。

② 此种方式往往是模型开发者在未经授权或超越授权的条件下,爬取数据持有企业的数据资源,进而涉嫌侵害数据持有企业的经济或商业利益。参见熊丙万、何娟:《数据确权:理路、方法与经济意义》,载《法学研究》2023年第3期,第60-62页。

③ 一般而言,依据数据持有者就企业数据的不同对外披露程度,可将其划分为不设置访问控制的公开数据(如社交媒体等平台已公开的信息)、设置访问控制的可(半)公开数据(如通过账号密码登录并访问、使用的数据库)与不公开数据(如商业秘密数据、其他重要竞争性利益的数据等)等三种类别。

④ 例如,《中华人民共和国数据安全法》第32条第1款规定,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。

⑤ See Jeffrey Neuburger, *hiQ and LinkedIn Reach Proposed Settlement in Landmark Scraping Case*, JD Supra, https://www.jdsupra.com/legalnews/hiq-and-linkedin-reach-proposed-1727186/, visited on 2025-2-1.

⑥ 参见上海知识产权法院(2016)沪73民终242号民事判决书。

数据在利用上仍然存在一定的不确定性——包括模型开发者在内的第三方使用者在收集与使用公开数据过程中的法律风险会进一步提高,持有企业投入的边际预防成本亦随之加大。

(四)“权益保护”与“技术创新”共存的治理难题

“法律必须稳定,但又不能静止不变。”[①]就数据领域“权益保护”与“技术创新”之间的关系而言,传统权益的保护是建立在过去的事实基础上,通过协调不同参与者间的利益冲突,来实现相对稳定的利益平衡状态。在大数据和人工智能时代,既有的权益保护规则似乎已经难以适应新兴技术的发展需求,模型开发者在数据收集、使用、内容生成等领域面临的法律风险,未能及时得到法律规则的回应。

从法律经济学的视角看来,之所以法律制度要保持稳定而不能朝令夕改,其原因不仅在于制度变动影响参与者对未来行动的预期,增加行为后果的不确定性,更在于制度变动的高昂成本以及相应风险。面对新的事实变化,原有的制度决策通常会产生相应的偏差,而纠正这种决策偏差也是有代价的——新的方案亦可能带来其他的不利影响或风险隐患。这就意味着,倘若纠正制度偏差的成本过于高昂,以至于超出了原有制度决策偏差带来的损失,那么保留这种偏差在经济上就是理性的。当然,这也并非意味着决策者只能固守着既有的治理规则,而对外界事物的发展与变化视而不见。鉴于因技术创新引发的制度变革所带来的预期收益评价的不确定性,理论界与实务界都在不同程度上寻求“制度变革成本”与“制度(保守)偏差损失”之间的平衡,以期通过合理的代价,最小化制度变革与制度保守可能带来的不利影响。犹如庞德在其《法律史解释》中所发现的那样,几乎每一次法学流派或理念的倡导或更迭,都是在为所处时代的实在法变革供给新的理论支撑;事后,一旦制度变革成为现实甚至出现矫枉过正之时,上述理论与思想又可能会成为抑制过度革新,回归与社会生活大体相适应的矫正性力量。[②]

总的来看,生成式人工智能的发展,进一步提高了数据在先权益保护与数据创新利用之间的协调成本。面对模型开发者对数据的规模化需求与创新利用需求,需要适合时代与技术发展变革需求的“回应型法”[③],来解决不同数据参与者之间利益冲突的治理挑战。但这种制度变革回应,既不是对原有权益保护随意添加额外的约束或限制条件,也不是对任意的技术创新需求在制度上给予放任性的满足,而是在权衡保守“权益保护”规则所带来的偏差损失与采取适合“技术创新”需求的制度变动可能带来的矫正成本之间,寻求一种新的利益平衡,使其能通过合理的代价,妥善解决因技术变革导致的既有数据权益保护与技术创新治理之间的紧张关系。这需要更为科学与充实的理据作为治理的支撑。

三、训练数据来源治理的理据支撑

在预设事实发生较大变动的前提下,如果将生成式人工智能训练数据的来源治理视为一种对

① [美]罗斯科·庞德:《法律史解释》,邓正来译,商务印书馆2016年版,第4页。

② 参见[美]罗斯科·庞德:《法律史解释》,邓正来译,商务印书馆2016年版,第4-6页。

③ 参见[美]诺内特、塞尔兹尼克:《转变中的法律与社会:迈向回应型法》,张志铭译,中国政法大学出版社2004年版,第85-87页。

传统制度的机械遵守与执行,那么在一定程度上会导致“规范”偏离“事实”。其结果,要么规范形同虚设,要么规范阻碍创新。如此一来,既有损于规范的权威,又会增加规范实施的负外部性。相反,在必要情形下,若将来源治理视为一种对原有制度的反思与矫正机制,则可在某种程度上促进“规范”与“事实”的有机融合,在合理限度内提升制度活力、弹性的同时,更有利于技术的创新与发展。对原有制度的反思与矫正,需要以更为有效的理据作为治理的支撑点,以增强变革原有制度的正当性。

(一)视角转换:损害的“相互性”考察

一般而言,损害的“相互性”源于科斯在其《社会成本问题》一文中,就侵害行为“单向度”分析的传统裁判思维给予的反思与批判。[①] 在“斯特奇斯诉布里奇曼案”(Sturges v. Bridgman)中,原告声称被告糖果厂在生产作业时发出的设备噪音影响到了其诊所(紧邻被告厨房)正常的诊疗服务,故向法院申请禁令,要求被告停止使用机器设备。[②] 事实上,在原告开设诊所之前,被告就已经在此处生产运营多年,若技术保持不变,被告开启使用机器设备就会产生相应的噪声并影响到周边的邻人(包括原告),除非让其追加额外的成本投入(购置隔音器材或搬迁到其他地方等),否则原告的诉求难以被满足。但若只考虑原告遭受的损失或负面影响,而忽视对被告可能面临的不利影响,很可能会导致既不公平亦非效率的后果。毕竟被告作为合法的厂商,从事生产经营活动原本就受到了法律的许可与保护。故在科斯看来,此案表面上可被视为一种“单向度”的侵权行为(被告的机器噪声侵扰了原告的正常执业),但实质上却是发生在原、被告之间的一种合法利益冲突[③],即不仅是被告的噪音对原告的执业活动产生了负外部性,原告的主张或诉求也可能会给被告的正常运营带来负外部性,例如原告申请禁令,要求被告停用机器设备等。对此案的裁判,就不宜单向、简单地认为仅仅是被告影响了原告的正当权益,因为这种思路很可能导致社会成本更大、社会收益更小的裁判后果,以至于造成司法激励的效率损失。若这种冲突无法在利益相关者之间通过合意解决,那么从成本收益的分析视角上观察,就要看哪一方给对方带来的“损害”更具效率。

倘若从权益损害的“相互性”视角来审视,在不考虑其他因素的条件下,既可将模型开发者对数据的利用视为对相关数据权利主体既有权益的潜在加害,也可把对数据在先权益的保护视为阻碍模型开发者技术创新的一种力量。换言之,这种保护也可能损害了模型开发者对数据创新的合理预期利益。这种损害的“相互性”,表明不同权益的行使或保障,都会给他人带来一定程度上的负外部性影响。由此可见,数据创新利用与数据在先权益——个人数据权益、作品数据版权以及企业数据权益的保障之间,往往呈现出一种“此消彼长”的关系。一般而言,当法律对数据在先权益保护水平越高,就意味着模型开发者对数据资源利用的自由空间更小,数据创新的可能性也就越低;反之,数据资源被利用的空间就越大,数据创新使用的可能性也就越高。

法律治理或干预的初衷不是以利益的先后作为保障的基础,而是在评价两种利益价值的基础上,尽量避免更大损害的出现,正所谓“两害相权取其轻”。模型开发者对数据的创新使用,在于挖

① See R. H. Coase, *The Problem of Social Cost*, 3 The Journal of Law and Economics 1, 2 (1960).

② See R. H. Coase, *The Problem of Social Cost*, 3 The Journal of Law and Economics 1, 8-10 (1960).

③ 暂不考虑政府或主管部门对厂商生产噪声的公共管制因素。

掘数据新的价值,追求的是形成新的价值增量;而对于既有数据权益的保护,则在于避免对已有权益带来损害。这意味着,只要模型开发者在数据创新利用中所带来的收益足够大,以至于在客观上补偿既有数据权益者的损失后仍存有剩余,那么相比于既有数据权益,开发者数据创新利用的价值就更大,也更具优先保护价值。理论上,这恰好也符合"卡尔多—希克斯"的效率改进标准(Kaldor-Hicks Principle)。①

当然,上述论断也只是一种可能,在实践中对既有数据权益与数据创新利用价值间的量化与比较,可能限于技术等多方面的因素影响,以至难以进行有效评价,例如模型开发者对数据的创新利用可能不仅仅具有收益的维度,亦可能因技术系统的不确定性而面临风险或隐患。若将这些影响因素一一加以分析,会使评价成本高不可攀。退而求其次的办法,则是尽量减少"数据权益保护"与"数据创新利用"相互间的损害影响,降低总体损失水平。在具体方案上,可将以上两者损害的"相互性"限定在一个合理的限度内,只需要避免给对方带来的超出合理必要限度的损害。

(二)利益平衡:最小化"权益保护"与"技术创新"的负外部性影响

在训练数据来源的治理中,面对"权益保护"与"技术创新"之间的数据利益冲突,单纯将某一方的权益予以静态的、教义式的固化,例如,将数据在先权益的保护视为某种不变的教义原则,或任意将技术的革新作为突破、超越已有规则的当然事由,均难以成为利益冲突的有效解决路径。与此相反,若将数据的权益保护与技术的创新利用放置于一个考量社会价值的动态框架下,那么,法律更有可能是在比较权益保护与技术创新的成本与收益基础上,选择在边际社会产出上更为有利的治理方案。

引申上述有关损害的"相互性"视角的分析,可以发现,按照既有治理方案解决模型开发者的数据来源问题,会与生成式人工智能海量数据的效率性获取需求相悖,同时在技术保持相对稳定的条件下,开发者亦难以通过合理的成本投入,来达到法律规范预设的相关要求,并最终会导致对技术创新的不当抑制。相反,若完全或近乎完全地放任技术创新的发展,虽可满足开发者在相关技术场景下对数据的自由利用,但又可能会因滥用等行为而对数据在先权益造成不当影响,进而在不同程度上减损数据创新利用的价值产出。随着生成式人工智能的技术发展,若要在数据权益者与模型开发者之间形成相对稳定的利益平衡状态,就需要克制彼此对他方利益带来的不合理负外部性影响。

首先,法律制度需要重新审视对数据在先权益的保护。以技术发展的合理需求作为权益保护水平的重要考量因素,并试图最小化对数据创新利用的不当影响,给数据创新留有足够的开发空间。否则,一味强调在先权益的保护,忽视潜在开发者对数据利用的各种创新机会,甚至以严厉的处罚措施威慑可能减损数据在先权益但又属于合理创新利用的行为,势必会影响数据以及未来人工智能技术的可持续发展。就此而言,欧盟目前未能出现在全球具有影响力的互联网企业或人工智能产品及服务,其原因之一就是欧盟在某些领域过于强调对数据权益的保护,对数据利用、人工智能开发的管制干预过于严苛,以至于大幅降低了本土企业或开发者数据创新利用的机会与可能。

① 参见[美]理查德·A·波斯纳:《正义/司法的经济学》,苏力译,中国政法大学出版社2002年版,第90-91页。

此外,以企业公开数据的利用为例,在 hiQ 实验室诉领英数据爬取纠纷案中,相对于原告而言,被告作为职场社交平台的运营者,对其持有的用户公开数据具有一种在先的管理与控制权利,但若对此种权利给予类似无限制、绝对性的排他保护,既与数据的非竞争性相冲突,又严重影响此类数据被第三方使用者以合理目的加以利用的可能。在当事人之间的损害权衡上,与被告的数据合作协议终止后,原告虽未经被告同意爬取其公开用户数据,但原告所从事的目的并非在于恶意损害被告商业利益或进行所谓的"同质化"商业模式,而是另行从事其他数据分析业务。此种活动既不会对被告的竞争利益产生不利影响,也没有证据表明会对用户权益造成损害。这至少意味着原告获取被告公开数据的负外部性并不明显,大体在合理的范围内。同时,若不允许原告收集、获取被告平台上的公开用户数据,原告很可能难以继续运营。由此可见,若其他条件不变,允许原告合理爬取被告公开数据,对被告新增的负外部性较小;而禁止原告爬取被告公开数据,则将对原告新增显著的负外部性。此外,从公共利益考量,允许原告爬取被告公开数据,并不会对社会利益造成更多不利影响。基于此,该案的初审法院(2017 年)①、上诉法院(2019 年、2022 年)②均支持了原告申请的初步禁令请求。尽管本案的原告并非生成式人工智能的模型开发者,但上述裁判似乎在某种程度上面向未来的第三方使用者(包括但不限于模型开发者等在内的数据挖掘主体)释放了一种可期待性的信号——与未经数据权利人许可不得收集利用的绝对保护模式相比,在一定条件下,允许第三方未经许可合理收集并使用企业的公开数据资源,更有利于此类数据的价值开发,且对于相关者利益冲突的协调,亦不失为一种可能且有效的解决方案。

其次,法律制度也要在一定程度上对技术的创新开发给予必要的克制。节制对权益的绝对保护,既不意味着是对数据在先权益保护的放弃或否定,也不意味着放任对技术的自由创新或应用可能产生的任何不利影响。通常情况下,技术的创新在提高生产效率与效能的同时,也会给相关领域带来不同程度的风险隐患,这不仅值得关注,更需要采取适当的方式加以有效治理。例如,针对美国 OpenAI 公司开发的大型语言模型 ChatGPT 等产品的问世,对隐私或个人信息、数据安全等方面带来的风险隐患或威胁,众多美国学界业界人士于 2023 年联名呼吁在一定期限内停止对 GPT 产品的研发,并加强对大模型及其产品在数据权益保护、安全性能等方面的评估。③ 与此同时,不少国家的监管部门亦要求暂停 ChatGPT 向社会的开放使用,并对其可能产生的潜在危害或风险进行必要评估。④

结合上述事例,倘若将创新过程与应用风险视为一种"黑箱",那么至少从目前的技术水平来看,完全打开这一"黑箱"的信息成本过于昂贵,以至于可能淹没由此产生的预期收益,但这并不意

① See hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099 (N.D. Cal. 2017).

② See hiQ Labs, Inc. v. LinkedIn Corp., 938 F.3d 985 (9th Cir. 2019); hiQ Labs, Inc. v. LinkedIn Corp., 31 F.4th 1180 (9th Cir. 2022).

③ See *Pause Giant AI Experiments: An Open Letter*, Future of Life, http://futureoflife.org/open-letter/pause-giant-ai-experiments/, visited on 2025-2-1. 当然,除技术风险之外,还可能包括一些主要呼吁人士(如埃隆·马斯克等)在其经济或商业上的利益考量。相关分析,亦可参见郭春镇:《生成式 AI 的融贯性法律治理——以生成式预训练模型(GPT)为例》,载《现代法学》2023 年第 3 期,第 96 页。

④ 参见泰勒:《突然宣布:"封杀"ChatGPT! 打响人类保卫战第一枪!》,载中国基金报官网,https://www.chnfund.com/article/AR202304010650061032749 4,2025 年 2 月 1 日访问;王进雨:《ChatGPT 遭遇封杀,监管要防止人工智能被滥用》,载新京报官网,https://www.bjnews.com.cn/detail/168060675814151.html,2025 年 2 月 1 日访问。

味着可以对技术创新的风险“黑箱”置之不理。相反,通过适当的权益保护与必要的合理性约束,或许会成为抑制或减少技术创新的负外部性、实现技术创新对社会福利最大化的一个可能的重要选项。《欧盟人工智能法》(EU AI Act)结合人工智能模型的风险等级因素,对相应模型的开发或创新活动进行了必要的管制。包括该法第5条规定的8项被禁止的人工智能实践①,以及第3章规定的“高风险人工智能系统”有关“合规要求”“风险管理系统”“数据和数据治理”“技术性文件”“记录保存”②等方面的限制性规定,均在不同程度上对模型及其数据训练等开发或创新活动进行了必要的约束,以避免或减少数据创新利用可能带来的社会风险及其损失。基于现代微观经济学视角,通常情形下,要使模型开发者保持技术创新的私人成本与社会成本大体一致,而非将大量的私人成本不合理地转嫁给第三方的主要激励,便是对开发者给第三方造成的不当负外部性予以内在化③,使其成为开发者技术创新活动必要且合理的约束,进而推动社会福利的增长。

总的来看,在生成式人工智能蓬勃发展的背景下,基于“相互性”动态关系模式,最小化数据在先权益的保护与数据创新利用的负外部性影响,或者说有效抑制数据权益的过度保护与技术创新的风险放任,更有利于在两者间形成新的利益“平衡点”。

(三)制度激励:合理空间的释放与权益界定的优化

“相互性”的视角,激发了“利益平衡”为主导的治理设想。但基于“利益平衡”的治理,终究都要通过具体的制度安排加以落实。在促进模型开发者与数据在先权益者之间利益平衡的基础上,对既有权益保护制度予以改进、优化,可降低前者在数据来源方面面临的法律风险,进而合理释放在生成式人工智能领域创新利用来源数据的更多空间。

在训练数据的来源层面,制度的改进与权益界定的优化,意味着要将模型开发者对训练数据的收集与使用需求纳入数据权益保护决策的考量范围。结合人工智能的技术发展规律及不同阶段的特点,矫正原有权益保护的决策偏差,适当限缩权益保护的范围,合理开放数据的创新利用空间,更有利于促进数字技术的发展。具体而言,对于已经在制度上形成相对固化“保护模式”的数据权益或相关权益,例如个人数据权益、作品数据中的版权等,原有的制度保护模式可通过合理使用机制拓展数据创新利用的空间。信息传播与利用是人类知识创新发展与演进的基本路径,模型开发者在合理限度收集、使用受保护的数据资源,不仅是对在先权益的尊重或价值认可,而且亦能促进此类数据的充分利用,激励额外的创新收益,有利于社会福利的增长。若无条件地将授权使用机制作为模型开发者数据收集、使用的刚性约束,则易不合理地限缩甚至是遏制数据资源创新利用的机会或潜在发展。将合理使用机制纳入数据在先权益的保护体系,通常也不会放任模型开发者对数据资源的任意收集或使用。倘若出现合理利用之外的情形,著作权人仍可基于保护规则寻求救济。

针对虽有保护倾向或初步规划,但数据在先权益的边界、不同数据参与者的行动范围尚不清晰的情形,由于这些权益边界、行动范围影响着企业公开数据的权益保护与开发利用事宜,故在此意

① 参见《欧盟人工智能法》(EU AI Act)第2章“禁止的人工智能实践”第5条第1款的规定。

② 参见《欧盟人工智能法》(EU AI Act)第3章“高风险人工智能系统”第2节的规定。

③ 参见[美]N. 格里高利·曼昆:《经济学原理:微观经济学分册》(第8版),梁小民、梁砾译,北京大学出版社2020年版,第205-206页。

义上,解决企业公开数据权益保护与开发利用之间的利益冲突,应当聚焦公开数据的权益界定。正如一条“机非混行”的道路,机动车与非机动车都有权在道路上行驶,但由于缺乏车道标线,致使二者发生交通事故(利益冲突)的概率通常会比“分道行驶”的道路更大。同理,就公开数据的保护与利用而言,无论是持有者对此类数据的在先管控利益,还是模型开发者(第三方使用者)对其的收集、使用利益,均为合法利益的范畴:若不考虑其他因素,前者的权益虽未在相应的制定法层面予以明确,但结合目前的政策文件等,企业对其合法收集、管控的数据资源,具有持有、使用以及产品经营等权益①;而对后者来说,至少是在民事领域,在不违背国家安全、公共安全或善良风俗等前提下,模型开发者自行收集、使用持有企业的公开数据,亦符合“法无禁止即可为”的原则。但正是由于此类数据在权益保护的范围与第三方开发利用的限度上,尚未得到较为清晰的边界划定,故二者发生利益冲突的可能性更大,且更易在实践中因大规模爬取公开数据而产生纠纷。在此基础上,对企业公开数据收集与使用的行为边界的划定,有利于减少因数据爬取与使用引发的纠纷,促进公开数据在必要管控与合理获取之间的利益均衡。鉴于此类情形在制定法层面尚未形成相对固化的“保护模式”,应当在必要时明确持有企业与模型开发者就公开数据资源的利益边界或行动权限,合理降低模型开发者在数据获取与使用上的法律风险。

总的来看,在生成式人工智能训练数据的治理层面,寻求数据权益保护与激励技术创新发展的制度平衡目标,并非是对既有权益保护的否定或放弃,也不是对数据创新利用持恣意或放任态度,而是秉持在权益保护与激励创新的治理中,施以合理且必要的约束或限制。由此,使得数据在先权益与开发者对数据的创新利用需求,能够在改进或优化后的制度框架下实现共存。

四、训练数据来源治理的方案探索

结合生成式人工智能的技术开发实践,探索模型开发者在个人数据与作品数据上的合理使用机制,可在一定程度上缓解其与数据在先权益者之间的利益冲突,拓展数据的开发利用空间,合理降低前者在训练数据来源的法律风险。另外,明晰企业公开数据的利益边界与保护模式,合理收集与使用互联网公开数据资源,亦有利于避免或减少此类数据开发利用的纠纷,为模型开发者获取与使用公开数据厘清合法路径。最后,通过嵌入符合数据创新利用需求的包容性与可行性的治理方式,更有助于降低上述革新方案的实施成本,增进治理的实效性。

(一)个人训练数据合理使用规范的构建

在本文语境下,个人训练数据是指用于生成式人工智能模型训练所需的个人数据资源。个人训练数据合理使用的机制设计,主要目标在于在兼顾既有个人数据权益的基础上,大致满足模型开发者数据训练的需要,以释放出更多的个人数据利用空间。在已有的制度安排框架下,已公开的个人数据可被纳入包括模型开发者在内的数据处理者合理使用的范围②,故个人训练数据合理使用规范的构建,需要更多关注未公开个人数据的收集与使用。

① 参见《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》(2022 年 12 月 2 日)第 2 部分。

② 参见《中华人民共和国个人信息保护法》第 27 条关于合法已公开的个人信息收集与使用的规定。

在数据收集的“告知同意”方面,原则上可构建非敏感个人数据的合理使用机制,以替代既有治理方案中事先取得同意的一般数据收集与利用规则,合理降低模型开发者在获取此类数据过程中的交易成本与法律风险。通常来说,以数据内容是否涉及个人的敏感信息而言,可将个人数据区分为敏感数据与非敏感数据。前者一旦被泄露或非法使用,导致危及数据主体人格利益或其他利益的风险会显著增加,这类数据通常表现为含有身份识别信息、宗教信仰、金融账户、医疗健康以及生物识别等方面的内容,且往往处于直接识别或已识别的状态。因此,无论收集使用者是否为模型开发者,都应当在法律保护框架下,履行相应的个人数据保护义务,诸如事前履行如实告知义务,并获得数据主体的同意等。但结合具体的应用场景,敏感个人数据的范围可能并不容易被把握①,故有必要依据不同场景与安全保护的程度对此类数据的外延进行严格限定,在有效保障个人数据权益的同时,也要避免敏感个人数据的泛化保护。

相对于敏感个人数据,非敏感个人数据往往涉及的信息价值有限,且在不少情形下,亦难以直接识别到特定自然人的身份,导致个人隐私泄露或相关权益被侵犯的可能性较小。故在严格限定敏感个人数据范围的基础上,就非敏感性个人数据而言,模型开发者原则上可不经个人授权或同意便直接加以收集或利用,但应表明相应的数据来源,并对数据处理的过程予以记录,以供主管部门或行业机构的监督检查或自律性审查。此外,法律亦不应强人所难。对于应当取得同意的个人数据,限于网络与数据处理技术等客观原因,模型开发者事前未经个人同意收集获得的,应当允许开发者事后采取补救措施,即在合理期限内取得个人同意,或予以删除、匿名化等处理,并在此过程中严格保护此类数据的安全与其他个人权益,以体现法律治理对技术创新实践的尊重与包容。

在个人数据处理的“目的必要性”方面,模型开发者在以模型训练为目的获取与使用个人数据时,对于获取数据在规模、限度上的合理性评价,应当在尊重模型开发技术规律的基础上,给予更多宽容,即除非有相反的证据证明此类数据被用于非法目的,原则上均可推定为获取与使用目的明确、合理。另外,为了防范个人数据处理者滥用其技术优势损害个人权益,亦可从增进此类数据在处理中的透明度视角,制约模型开发者对个人训练数据收集与利用中的“黑箱”操作。具体而言,可要求模型开发者自行制定个人训练数据收集、使用规范,报主管部门或行业机构备案并面向社会公开,以增强模型开发者对此类数据处理情况的信息披露,减少其与数据主体在个人数据处理上的信息不对称。在此基础上,模型开发者亦应设置合理的个人训练数据使用异议机制,主动接受包括数据主体等在内的社会各界的监督与信息反馈。

值得一提的是,鉴于生成式人工智能的发展,作为大模型内容输出的合成数据已被业界广泛关注。② 在应用领域,利用合成数据的“深度仿真”等功能,可通过虚拟场景模拟复杂、获取代价昂贵的现实场景。理论上,可以利用合成数据来模拟甚至是替代个人数据,以减少模型训练收集、获取真实数据的成本,并降低隐私泄露等数据事故的概率。③ 事实上,此种替代机制确实能够增强对个人

① 参见郭传凯:《敏感个人信息处理规则的反思与修正》,载《政法论坛》2024年第3期,第102-113页。

② 参见曹建峰、陈楚仪:《AIGC浪潮下,合成数据关乎人工智能的未来》,载《新经济导刊》2022年第4期,第25-31页。

③ 参见刘培:《人工智能合成数据:伦理隐忧与风险治理》,载中国社会科学网,https://www.cssn.cn/skgz/bwyc/202410/t20241029_5797221.shtml,2025年2月1日访问。

数据权益——特别是敏感数据的技术性保护,利用虚拟数据来模仿真实数据的个性化特征,进而减少数据处理事故的发生。但合成数据并非凭空而来,生成或创制合成数据仍为基于算法的数据分析处理的产物或附随品,前提自然不能缺少真实的数据资源,这也就意味着模型开发者仍然需要获取真实的个人数据作为合成数据生成或创制的基础。况且,合成数据毕竟是模拟真实数据的产物,其与真实数据仍可能存在不同程度的偏差,大规模的合成数据资源,势必也会融入更多人为(技术人员)的主观意志甚至是偏见,进而会影响到模型内容的输出结果。故以合成数据替代真实个人数据的解决方案,或许需要更多的经验积累与技术试验,方可获得较为客观的结果。当然,在现有技术水平下,可将合成数据作为真实个人训练数据的一种尝试性、试验性的替代机制,并持续跟踪、评估此种替代机制对模型训练质量与个人权益保护可能产生的利弊影响。

(二)作品训练数据合理使用范围的拓展

与作品的传统利用方式不同,开发者对作品训练数据的利用在于满足模型机器学习与深度学习的需求,即通过一定的算法对其进行挖掘分析处理,而非直接对此类数据在内容或表达上的欣赏,或者简单且机械地复制、模仿。就作品用于数据训练这一问题,学界大致有三种不同意见。第一种意见认为,作品数据用于模型训练受版权法调整,但可通过合理使用机制来创造模型对训练数据集的广泛获取、合理学习的机会①;或鉴于其复制权、信息网络传播权被侵犯的风险依然较大,原则上仍以授权使用模式作为主导,但可构建多元化的授权机制作为训练数据供给的合理渠道②;第二种意见认为,应当结合不同的利用场景,区分用于机器学习的作品数据是否属于版权法保护的范围,将特定作品的表达性使用纳入版权法保护的范围,将非表达性使用作品的行为排除在版权法保护的范围之外③;第三种意见认为数据训练对作品的使用具有明显的“非特定性”④,故应排除版权法保护的范围,并另行界定为“非表达性使用”的行为。⑤

结合上述分析,若仍按照既有版权作品授权使用的方案,开发者获取用于模型训练的作品数据的交易成本会进一步提高。根据需求规律,这也就意味着开发者可能在一定程度上会减少对此类数据的需求量,进而影响训练数据或模型输出的质量,同时也会大幅提高未来数据创新利用的成本;若完全将模型训练对作品数据的收集利用视为一种“非表达性使用”,排除版权法的保护,则亦与现实不符。在学习方法层面,尽管生成式人工智能的机器学习着重掌握文字间、图像元素间等的分布规律,而非是对作品特定性、表达性的欣赏或复制;但机器学习的结果最终仍是用于输出高质量的内容表达,尽管这种结果是通过数据形式加以表现的。无论是从符号意义上分析语言(规律),还是从内容意义上表述语言,相对于劣质的输入内容,好的内容(作品)在经过机器挖掘分析后,通常会产生更多更为优质的输出内容。在此意义上,作者对于模型训练输出的高品质内容,仍具有一定的贡献。只不过这种贡献度相较于传统的作品(直接)使用而言有所减弱,即在通常情况下,从传

① See Mark A. Lemley & Bryan Casey, *Fair Learning*, 99 Texas Law Review 743, 745-748(2021).

② 参见张平:《人工智能生成内容著作权合法性的制度难题及其解决路径》,载《法律科学(西北政法大学学报)》2024 年第 3 期,第 28-31 页。

③ 参见李安:《机器学习的版权规则:历史启示与当代方案》,载《环球法律评论》2023 年第 6 期,第 97 页。

④ 参见刘晓春:《生成式人工智能数据训练中的“非作品性使用”及其合法性证成》,载《法学论坛》2024 年第 3 期,第 68 页。

⑤ 参见陶乾:《基础模型训练的著作权问题:理论澄清与规则适用》,载《政法论坛》2024 年第 5 期,第 162 页。

统的对特定作品的直接使用,转向在某种程度上以获取特定内容为目的的文本或数据分析行为。这就意味着,用于模型训练的作品数据,并非完全不顾作品的“特定性”因素,而是适当拓展了作品“特定性”的范围,并针对性地提取作品的表达加以学习。故一般而言,以生成式人工智能模型训练为目的,在采取相应安全保障的前提下,将模型训练所需的作品数据纳入版权客体合理使用的范围,亦不失正当性与可行性。

鉴于生成式人工智能所需的训练数据规模庞大,特别是对于原本就属于相对稀缺的作品资源,其对开发者的模型生成与优化显得更为重要。故将用于模型训练的作品数据纳入版权法的合理使用制度,通常并不妨碍作者对其版权权利的行使,同时亦能在某种程度上“威慑”开发者对作品数据的滥用。另外,即便是将作品用于预处理或模型训练时作品被重组、聚合等,其目的亦仅在于机器学习与分析处理,提高模型内容生成的质量,而非有意损害作品的完整性。

在域外的立法实践上,日本与欧盟的版权法有关合理使用制度的创新,为人工智能时代对高品质训练数据的需求开拓了一条更为宽广的获取渠道。例如,日本在2018年的版权法修改中,将“不以欣赏作品所表达的思想或感受为目的的使用”作为版权“柔性”合理使用的一种新型方式,以此回应人工智能时代对大数据挖掘分析等技术创新的迫切需求。① 2019年通过的欧盟《单一数字市场版权与相关权指令》(Directive on Copyright and Related Rights in the Digital Single Market)也将科研机构与文化遗产机构满足相关条件的“以科学研究为目的的文本和数据挖掘”等行为,纳入版权合理使用的范围。②

结合上述分析,可将作品训练数据的合理使用规则进行如下的构造:一方面,原则上,模型开发者以生成式人工智能模型训练为目的,获取与使用作品无须经著作权人许可。另一方面,为了平衡与著作权人的利益,模型开发者在收集、使用作品数据用于模型训练时,应受合理性限度的约束,不得妨害著作权人对其作品的权利行使,同时亦不应给著作权人增添不当的额外负担。鉴于作者对其作品的生成投入了创新性的精力,任意无偿使用此类数据,难免会在一定程度上减损对内容生产与创新的激励。更何况,模型开发者从事的数据训练活动,主要以商业利益为目的,故对其的获取与使用理应体现出必要的合理约束,以防范对此类数据开发的滥用。

对此类数据用于模型训练的合理性约束,应当考虑作品来源的合法性因素。2025年6月23日,针对一起收集、利用作品数据“投喂”模型训练的案件,美国加利福尼亚北区联邦地区法院作出的初审裁判,或许可供借鉴。③ 在该案中,原告(图书作者)指控被告未经许可获取、使用其图书资源用于生成式人工智能的模型训练。法官经审理后认为,模型开发者将具有版权的作品用于数据训练与模型开发的行为,属于版权法上的合理使用;但为避免开发者对版权作品的滥用,应当对此类数据在来源上进行合法性控制。也就是说,开发者对用于模型训练的作品内容,无论是将非数字化的作品(如书籍)转化为数字化形式,还是直接获得相关作品的数据,均应当确保其来源合法。总的

① 参见郑重:《日本著作权法柔性合理使用条款及其启示》,载《知识产权》2022年第1期,第112-130页。

② 参见谢新洲、朱垚颖:《信息资源管理视角下的欧盟数字版权保护研究》,载《信息资源管理学报》2020年第6期,第61-69页。但也应当看到,该立法例对于纳入版权法合理使用的数据训练范围较窄,似乎难以满足广大模型开发企业的发展需求。

③ See Andrea Bartz, Charles Graeber, and Kirk Wallace Johnson *v.* Anthropic PBC, No. 3:24-cv-05417 (N. D. Cal.).

来看,通过作品源头获取的合法性控制,一方面,可保障版权人的正当利益不受非法侵犯,进而激励开发者通过合法渠道获得作品内容;另一方面,则可使作品内容的品质不受非法形式(如盗版者对原作品内容的篡改、删除等)的干扰,有利于高质量训练数据的供给。

(三)企业训练数据权益界定的优化

本文所谓的企业训练数据,是指用于生成式人工智能模型训练所需的各类企业数据资源。① 在现阶段,模型开发者的训练数据在很大程度上依赖于企业数据的供给。而数据权益界定的得当与否,又往往影响着数据资源在市场中的优化配置,至少是为数据资源的市场配置提供一个初始起点——数据权益的边界及其相应的保护模式。鉴于持有企业对可(半)公开数据、非公开数据的持有管控权益以及授权使用保护模式,已在立法与司法实践中得以清晰展现,故优化企业数据的权益界定则应主要聚焦于公开数据的权益划分与保护模式上,并为此类数据后续的交易、流通厘清合法性基础。

为此,除涉及公共利益、人格利益等因素外,企业公开数据的利益相关方应主要聚焦于持有企业与第三方使用者。如前文所述,通常情况下,持有企业对公开数据的“管控利益”与第三方使用者对此类数据的“使用利益”,均具合法性,只不过这两种合法利益的行使影响了相对方,以至于给相对方的利益行使带来一定的阻碍,也即权益行使的负外部性影响。在权益的初始配置上,相对于第三方使用者,持有企业往往难以通过合理的成本或代价,来有效认知或处理来自外界获取、使用数据的目的、规模以及方式等方面的复杂性与多样性。为了避免利益冲突解决可能产生的更多社会成本,将公开数据的权益赋予持有企业,而由信息更为充分、且能以较小成本避免数据处理事故的第三方使用者承担预防责任,更具经济效率。

在权益的保护方式上,若采取以“授权使用”的财产规则模式,那就意味着将公开数据的定价权限也赋予了持有企业一方,并使得其在信息、技术以及定价权限上都占据了足够的优势,第三方使用者则处于明显的劣势地位。此种模式较易引发持有企业的机会主义行为,并可能不合理地限缩第三方使用者对公开数据的利用空间。为了矫正这种在信息、技术以及定价上的不对称性,通过责任规则的方式对其加以保护,或许更为公平、合理:即允许第三方使用者在不经持有企业的许可下利用公开数据,但需支付对价,以作为对持有企业的补偿。通常情况下,这种对价往往是由交易者之外的公共权威加以确定,且在对价的形式上亦可施以“合理的约束条件”作为有偿利用的替代,即法律可对模型开发者在内的第三方使用者,在数据获取的目的、规模以及方式上进行必要的限制,作为该方利用持有企业公开数据正当性的关键评价因素。例如,可以“创新性利用的目的”“与处理目的相适应的获取数量”以及“不妨害持有企业合理设置的技术措施或规范”等标准作为开发者在数据获取目的、规模以及方式上的必要限度。

由此可见,可以“第三方使用者无害使用”作为模型开发者利用公开数据的原则,即对于未设置访问权限的企业公开数据,在不妨害公共利益和数据安全的情况下,模型开发者可不经持有企业许

① 企业训练数据与个人或作品训练数据有时存在重合之处,例如企业持有的个人数据,若用于模型训练,则此类数据既是企业训练数据,又是个人训练数据。在此基础上,模型开发者应当分别按照企业训练数据、个人训练数据以及作品训练数据的相关制度安排,收集与使用上述数据资源。

可,但应在数据获取目的、规模以及方式等方面满足相应的要求;若给持有企业增加额外不合理成本或损失的,应当给予相应补偿。

(四)治理方式包容性与可行性的提升

值得注意的是,上述方案的实施效果仍面临诸多挑战。诸如,在生成式人工智能模型开发过程中,限缩训练数据在先权益的保护范围,是否会造成更多的涉及个人数据权益或作品版权的利益纠纷;企业公开数据权益的优化配置,在实践中是否会导致持有企业降低对公开数据的社会供给,或开始考虑转变这种商业模式;上述方案能否提高开发者对数据的创新利用效率,解决来源数据在收集与使用上的困境等等。因此,对训练数据来源的治理,不能仅仅考量这种治理可能产生的预期收益,更要着眼于新的制度可能引发的其他负面影响。在实践中,正如"科林格里奇困境"(Collingridge Dilemma)所预示的那样,新制度的负面影响,往往难以在事前做出较为准确的评估,主要原因或许在于现有人工智能技术的发展变化过快,而法律制度及其变革往往会滞后于技术的快速发展①,并且技术蕴藏的风险或不利影响还未得到更为全面、有效的把握。

直面法律的滞后性,也并不意味着一定要以"迅雷不及掩耳之势"出台刚性治理措施,以强制手段预防和控制所谓的技术风险和隐患。安全与发展、保护与创新等理念始终是人工智能法律治理所面临的重要难题。② 但受到现有技术水平以及信息不充分性等因素的制约,那种刚性且直接的治理方式,很可能会提高法律干预偏差的概率,增加社会损失。倘若这种社会损失超过制度变革的预期收益,那么法律暂时的不干预就是一种理性的选择。从长远角度来看,法律暂时不干预是为了今后更好地干预。为了有效降低新制度可能产生的不利影响或风险,应当提高变革后制度实施的可行性,并逐步构建并发展出一套动态的、以柔性指引为主导的治理模式。

法律的生命在于实施。在生成式人工智能模型开发的场景下,这种动态、柔性的治理模式,不仅是一种对新技术、新业态包容治理的机制③,更是降低制度变革后实施成本、提高制度可行性水平的公共治理策略。这种动态、柔性的治理模式,在创制目标上强调对新兴技术及其开发运用的包容性,通过弹性的制度变革安排,评估对技术创新与权益保护带来的实际效果与影响。例如,在约束条件下,结合实际场景设置制度实施的观察期、试验期,并在此期间观察或评估制度实施对各方数据参与者的利益影响,而暂不进行额外管制性干预,或延迟管制性干预的时间等。在具体的措施层面,通过避风港规则、促进性方案或指南等方式,建立制度实施的柔性指引机制,有效降低包括模型开发者在内的数据参与者在技术创新过程中或制度变动后的法律风险。例如,避风港规则可激励符合条件的模型开发者自愿"进港"从事研发活动,就用于模型训练的个人数据、作品数据以及企业数据等收集与使用上获得相应法律责任豁免④;促进性方案或指南,则强调监管部门结合不同的场景,出台相应的促进性规则或指南,结合制度的变动情况,告知模型开发者在来源数据收集与使用过程中可能面临的法律风险,以及收集与使用来源数据的可行性评价标准,以期不断改进与优化制

① 参见梅夏英:《复杂系统与智能涌现:未来数字法研究的范式图景》,载《法学家》2024年第5期,第49页。

② 参见徐磊:《发展与安全并重:生成式人工智能风险的包容审慎监管》,载《理论与改革》2024年第4期,第67-69页。

③ 参见谢新水:《包容审慎:第四次工业革命背景下新经济业态的行政监管策略》,载《西北大学学报(哲学社会科学版)》2019年第3期,第37-39页。

④ 参见戴昕:《作为法律技术的安全港规则:原理与前景》,载《法学家》2023年第2期,第42-46页。

度实施的可行性。

五、结语

数字科技的发展与变革,需要具有与之匹配的制度方案,并予以必要的法律保障。面对生成式人工智能训练数据在来源层面的场景风险及其治理挑战,需要通过协调、改进或优化个人数据权益、作品数据版权以及企业数据权益等领域的既有制度安排,从而最小化数据权益保护与技术创新的负外部性影响。在合理保护数据既有权益的同时,法律亦应当为模型开发者对来源数据的获取提供必要的可行路径,同时明确训练数据在后续开发处理与流通利用等环节的合法框架。JS

Governance of Training Data Sources for Generative Artificial Intelligence

GU Chuan

(Beijing Water Resources Comprehensive Law Enforcement Corps, Beijing 100036, China)

Abstract: Training data constitutes one of the key elements in the construction and optimization of generative artificial intelligence (AI) models, serving as the fundamental prerequisite for ensuring the quality and reliability of AI-generated content. In light of the growing demand for large-scale and diversified data in model development, it is essential to address the potential legal risks associated with training data from different sources. Going forward, it is necessary to minimize the negative externalities arising from the tension between data rights protection and innovative utilization by coordinating, improving, and optimizing existing legal arrangements concerning personal data rights, copyright in creative works, and enterprise data interests, in order to achieve a new balance between model development and rights protection. While safeguarding existing data rights, the law should also enable a fair and reasonable expansion of data resource utilization in model construction and optimization, and provide feasible and legitimate pathways for developers to obtain source data.

Key words: generative artificial intelligence; training data; legal governance; law and economics

本文责任编辑:常烨　廖吕有