人工智能模型数据的预防性法律规制

胡锦璐[1,2],邱丽玲[1,2]
(1.广西警察学院法学院,广西南宁　530029;2.广西警察学院警察法学研究中心,广西南宁　530029)

[摘　要]人工智能产品的潜在风险主要来源于模型数据存在缺陷而导致模型输出结果缺陷,因此从模型数据治理的角度进行预防性法律规制,可以从源头上控制人工智能模型输出结果的正当性及安全性。应当将人工智能模型数据纳入国家标准体系进行分级分类管理,规范数据构建的过程和方法,明确数据提供者、开发者及使用者的法律责任边界。从责任分配的角度构建法律机制,迫使相关方在数据构建和使用过程中保持谨慎,防止由于数据缺陷导致的输出结果不合理、违法或不安全的情形,避免数据缺陷造成的损害后果。针对不同的人工智能产品制定数据获取、使用和分类分级的国家标准和行业标准,可采用无限制、限制、特许、禁止等使用级别,并设立专门机构对人工智能产业进行监督管理。

[关键词]人工智能;数据缺陷;法律风险防范;责任分配;监督管理
[中图分类号]G71　　[文献标识码]A　　[文章编号]2096-711X(2025)11-0176-03
doi:10.3969/j.issn.2096-711X.2025.11.059　　[本刊网址]http://www.hbxb.net

引言

人工智能产品和其行为所带来的数据安全、隐私保护、知识产权纠纷、算法偏见等多方面的法律风险、应当以何种模式来承担法律责任以及今后应当如何对风险进行有效控制成了主要讨论的内容。应当在基础层面寻找问题原因并提出解决方案,而模型数据作为人工智能模型最基础的“加工原料”和决定性要素应当成为最重要的研究对象。

一、人工智能产品潜在风险的核心来源:模型数据的缺陷

数据作为人工智能模型训练和运行的基础,其质量直接决定了模型输出结果的准确性和合理性。如果数据存在缺陷,模型的错误决策将带来法律风险。目前人工智能技术的发展水平仍然处于起始阶段,无论从人工智能产品还是人工智能行为上看依然存在不少的瑕疵而引发风险和争议。包括侵权责任风险、数据和信息安全风险、人工智能生成内容归属权争议、人工智能刑事犯罪风险、人工智能算法偏见风险、消费者保护问题以及人工智能“奇点”风险。

数据的缺陷是导致人工智能模型风险的主要原因。数据、算法、算力共同构成了人工智能的三大核心要素,算法是基于训练模型数据的基础上遵循特定数学逻辑而生成结果的结构化决策过程。模型数据是人工智能算法深度学习、分析和决策的对象,是经算法深度学习后所合成结果的基础性要素,也直接决定了最终输出结果的实质内容和法律效果。

人工智能侵权风险的类型上,主要存在数据采集侵权和输出结果侵权,数据采集侵权主要由于采集数据信息来源的不合法或未经授权导致了使用数据不合法,从而进一步导致输出结果一并处于侵权状态。数据和信息安全风险与模型数据瑕疵之间的因果关系则更为直接。标注数据质量不佳发生成毒害内容、预训练数据和测试数据代表性不足导致价值观偏差、敏感数据使用和保管不当导致信息安全泄露等数据和信息安全风险,直接或间接原因都来源于模型数据瑕疵所形成的输出结果和行为效果瑕疵风险。人工智能刑事犯罪风险包括模型数据瑕疵和算法偏见导致的严重侵犯合法权益而触及刑事责任的类型,如煽动教唆类犯罪、侵犯知识产权类犯罪,发掘其输出结果的来源,都直接或间接地因模型数据瑕疵而起始。人工智能算法偏见风险成因之一也是模型数据偏差的瑕疵所导致的模型输出结果瑕疵,日常生活中的数据包含的人类社会固有的偏见和开发者本身主观和客观所具有的偏见都难以避免地在采集和模型训练的时候成为算法的一部分被人工智能所采用和输出。

二、人工智能模型数据的法律规制的责任分配模式

从人工智能模型运行的原理和结构来看,基础就是模型数据,若要对人工智能从根源上进行法律治理就必须首先对模型数据进行规范治理。然而人工智能抓取和学习的数据是海量且不确定的,要想分别逐一对数据进行识别和规范实无可能,必须寻找一个能够对海量数据进行梳理和鉴别的方法。

法律对大模型数据行为的有效规制,对于防止人工智能技术所引发的各种风险具有重要意义和实践可操作性。第一,通过对数据的来源合法性和质量要求的控制,可以防止数据的非法收集和使用,同时避免因数据不准确或有偏差导致模型决策失误。第二,法律要求对存储和传输的数据采取严格的安全措施,可以防止数据泄露和非法访问,同时建立数据泄露应急响应机制,也可确保在发生数据泄露时能够及时应对减少损失。第三,法律对大模型数据行为的规范作用,可以提高数据透明度,保障用户的知情权,确保数据来源和使用过程可追溯。

在数据采用原则上,可以“法不禁止即可为”作为原则,即只要是不受限的数据信息即可自由采用,只对相应受限制和禁止的数据进行清理。同时应制定一套完整的数据分级分类标准,只需对特定类型的特定级别的数据进行管理和规范,从数据治理的维度上对数据行为进行法律责任的规范和分配,明确模型数据各相关主体在数据行为或输出结果中应当承担的责任和义务。

第一,在数据收集阶段,参与数据行为的主体包括数据提供者和模型开发者。对于数据提供者,在提供数据时应当向数据主体说明数据将如何被收集和使用,同时获得数据主

收稿日期:2024-11-12
基金项目:本文系广西壮族自治区公安厅专项课题“少数民族地区边防检查行政执法规范化研究”(项目编号:2023GAQN078)。
作者简介:胡锦璐(1991—),女,广西柳城人,广西警察学院法学院助理研究员,广西警察学院警察法学研究中心研究人员,法学博士,研究方向:行政法学。

体的明确同意并告知数据的用途、处理方式和存储期限。对于模型开发者,在获取和使用数据之前,应当验证数据提供者的合规性,确保数据收集的每个步骤都符合相关法律和伦理标准。

第二,在数据处理和存储阶段,参与数据行为的主体包括数据处理者和数据储存者。对于数据处理者,处理数据时应当采取必要的技术和组织措施保护数据安全,防止数据泄露、篡改和未经授权的访问,应当遵循数据最小化原则,只处理为实现特定目的所必需的数据。对于数据存储者,应当确保数据存储系统安全可靠,数据存储符合相关法规要求,如数据加密、访问控制等,应当遵守数据保留规则,确保在法律规定的期限内存储数据,并在不再需要时安全销毁数据。

第三,在数据使用阶段,参与数据行为的主体包括模型开发者和模型运营者。对于模型开发者,在训练和使用AI模型时,应当确保数据的使用符合预期目的,并避免对数据主体造成不必要的隐私侵害,并应提供有关模型训练和数据使用的透明度报告,向利益相关者说明数据如何用于AI模型的训练和推理,以保证使用的数据不会导致模型产生偏见或歧视,采取措施减少算法偏见。对于模型运营者,在部署和运营AI模型时,应当确保其行为符合相关法律法规,特别是在数据隐私和安全方面。

第四,在数据共享和转移阶段,参与数据行为的主体包括数据共享者和数据接收者。对于数据共享者,在数据共享过程中,应当确保数据接收方符合相关法律要求,并签订数据保护协议,使数据隐私和安全得到有效保护,不会因数据转移而引发安全风险。对于数据接收者,在接收数据后,应当遵守数据保护法律法规,确保数据的合法使用和处理。

通过明确在数据收集、处理、存储、使用和共享各个阶段的法律责任,可以有效防止人工智能大模型技术引发的风险。每个相关方在其角色范围内承担相应的法律责任,确保数据处理的合法性和合规性,保护数据主体的隐私和权益。这种系统性的责任分配机制有助于建立透明、公正和负责任的数据生态系统。

三、对模型数据预防性法律规制的路径

我们可以在模型数据获取、使用和分类分级等方面建立一套全面的管理体系和法律体系,确保人工智能大模型技术在开发和应用过程中合法、合规,保护数据主体的权益,促进人工智能技术的健康和可持续发展。

(一)对模型数据行为分阶段分主体进行法律规制

人工智能大模型在数据行为方面的预防性法律规制,应当在数据获取、处理、存储和使用的各个阶段,需要系统性地设计法律措施,以预防潜在的法律和伦理风险。模型数据的标准化是人工智能法律治理的必然过程,也是对人工智能进行法律治理的前提条件。

在数据获取阶段,法律应明确数据来源的合法性和透明度要求。监管机构在此阶段的职责是建立数据收集审查和备案制度,数据提供者在收集数据前需向监管机构备案,并接受审查。在数据处理和存储阶段,法律需规定数据处理合规性和数据安全保障措施。监管机构的职责包括定期对数据处理者进行审计,检查其数据处理活动是否符合法律要求,并在发生数据泄露或其他安全事故时,要求数据处理者立即向监管机构报告,并采取相应的补救措施。在数据使用和共享管理方面,法律需明确规定数据使用的用途限定和透明度要求。监管机构的职责包括对数据使用者的数据使用活动进行监督,确保其符合法律和伦理标准,并对跨境数据流动进行审查,确保其符合相关法规,保护数据主体的权益。

在数据处理的各个环节,法律应明确各方的法律责任,确保在出现问题时有明确的责任主体。对于多方共同参与的数据处理活动,规定各方的连带责任,确保数据主体能够获得充分的救济。对于违反数据保护法律法规的行为,规定相应的行政处罚措施,如罚款、禁令等;对于严重侵犯数据主体权益的行为,追究相应的刑事责任。监管机构在此阶段的职责包括负责法律的执行,对违法行为进行处罚,确保法律的权威性和有效性,并设立数据主体投诉渠道,接受并处理数据主体的投诉和申诉,提供有效的法律救济途径。

(二)模型数据分类分级标准纳入国家标准体系

我国针对人工智能的分类分级目前暂时没有具体规定。在数据分类和分级管理方面,法律应明确数据分类和分级标准。应明确区分敏感数据和普通数据,对敏感数据进行特别保护,并根据数据的敏感性、重要性和使用风险,制定数据分级标准。针对不同级别的数据,制定相应的处理、存储和访问控制措施。监管机构在此阶段的职责包括制定并发布数据分类和分级的指导文件,帮助数据处理者正确分类和保护数据,并对数据处理者的数据分类和分级情况进行检查,确保其遵循相关法律规定。

数据分级可采用无限制、限制、特许、禁止等使用级别。数据在无限制级别下,可以被广泛使用,但仍需符合相关法规和道德规范,应当注意防范滥用风险,确保数据的隐私和安全。数据在限制级别下,具有一定的使用限制。可能限制数据的使用目的、使用者身份、使用时间等方面,限制级别可用于保护敏感信息和确保合规性。数据在特许级别下,用户必须获得授权或许可才能使用,特许级别可能应用于商业化的数据产品,需要购买许可或签署合同后才能合法使用。数据在禁止级别下,明确禁止任何形式的使用。数据可能包含高度敏感信息,不得用于任何目的,用户将不得使用数据。

(三)完善人工智能产业监管机构职能和教育培训体系

设立专门机构对人工智能产业进行管理,可以引导产业发展方向,制定战略规划和政策,引导人工智能产业朝着创新、可持续和社会有益的方向发展。设立专门机构还可负责制定合适的法规和标准,包括数据隐私、伦理规范、技术标准等,以保障公众利益、维护企业合规性,负责监督和管理人工智能产业的合规性,降低潜在风险。同时,监管机构也应肩负起引导人工智能产业朝着正确的方向发展的责任,应当完善对产业相关主体的教育培训体系,使相关产业主体接受数据保护法律法规的培训,增强其法律意识和合规能力。

结语

人工智能数据法律规范化能够实现数据的自由流动和创新发展,是应对人工智能带来的伦理、社会、经济等挑战的必要措施,也是一个长期的、动态的、开放的过程。实现人工智能数据法律规范化,能够平衡数据的私有性和公共性,协调数据的竞争和合作,解决数据的利益冲突和分歧,形成数据的共治共享机制。

参考文献:

[1]王利明.生成式人工智能侵权的法律应对[J].中国应用法学,2023(5):27-38.

[2]张欣.生成式人工智能的数据风险与治理路径[J].法律科学(西北政法大学学报),2023,41(5):42-54.

[3]支振锋.生成式人工智能大模型的信息内容治理[J].政法论坛,2023,41(4):34-48.

[4]吴冠军.通用人工智能:是"赋能"还是"危险"[J].人民论坛,2023(5):48-52.

(下转第180页)

政课教学面临的困境与推进策略的进行探索,不仅能够为思政课教学注入新的活力与可能,进一步提升思政课的吸引力和感染力,还能够为培养德智体美劳全面发展的社会主义建设者和接班人贡献力量。

参考文献:

[1]甘子成. 数字化红色文化资源赋能高职思政课教学改革创新探究[J]. 中学政治教学参考,2022(7):70-72.

[2]彭庆红. 善用数字技术建好"大思政课"[J]. 中国高育等教育,2024(9):49-54.

[3]陈亚红. 红色文化数字化赋能高校思政课研究[J]. 盐城工学院学报(社会科学版),2023(6):96-99.

[4]庞春阳. 红色文化资源数字化赋能高校思政课论析[J]. 高校马克思主义理论教育研究,2023(4):113-118.

[5]田珊. 数字化红色文化资源赋能高校思政课的价值及路径探析[J]. 思想理论教育导刊,2022(7):155-159.

Research on the Challenges, Strategies and Effects of Empowering Ideological and Political Education in Universities with Digital Local Red Cultural Resources

QIN Yan-chun

(Baise Vocational College, Baise Guangxi 533000, China)

Abstract: This paper delves into the challenges encountered in leveraging digital local red cultural resources to empower the teaching of ideological and political courses (IPCs) in universities, focusing on aspects such as the application of digital technology, resource collection and integration, innovation in teaching modes, and student engagement. To address these challenges, based on the strategy of "Four Modernizations and One Enhancement", the following strategies are proposed: strengthening the construction of the teaching faculty and enhancing their digital teaching skills; integrating local red cultural resources to build a digital teaching platform; innovating teaching modes and methods to improve the effectiveness of IPC teaching; and establishing a digital evaluation system to ensure steady improvement in teaching quality. These strategies have effectively elevated the teaching effectiveness of IPCs, providing robust support for the innovation and development of IPC teaching.

Key words: digitalization; local red cultural resources; teaching of ideological and political courses (IPC teaching)

(责任编辑:杨雨青)

(上接第 177 页)

[5]陈兵,董思琰. 分类分级治理算法的基本内涵及实践进路[J]. 西安财经大学学报,2023,36(6):70-79.

[6]张凌寒. 深度合成治理的逻辑更新与体系迭代——ChatGPT 等生成型人工智能治理的中国路径[J]. 法律科学(西北政法大学学报),2023,41(3):38-51.

[7]刘艳红. 生成式人工智能的三大安全风险及法律规制——以 ChatGPT 为例[J]. 东方法学,2023(4):29-43.

[8]刘宪权. ChatGPT 等生成式人工智能的刑事责任问题研究[J]. 现代法学,2023,45(4):110-125.

[9]孟令宇. 从算法偏见到算法歧视:算法歧视的责任问题探究[J]. 东北大学学报(社会科学版),2022,24(1):1-9.

Preventive Legal Regulation of Data Used by Artificial Intelligence Models

HU Jin-lu[1,2], QIU Li-ling[1,2]

(1. Law School, Guangxi Police College, Nanning Guangxi 530029;
2. Police Legal Research Center, Guangxi Police College, Nanning Guangxi 530029, China)

Abstract: The potential risks of artificial intelligence products mainly come from the defects in the model data that lead to defects in the model output results, so preventive legal regulation from the perspective of model data governance can control the legitimacy and security of the model output results from the source. The model data of artificial intelligence models should be incorporated into the national standard system for grading and classification management, standardizing the process and methods of data construction, and clarifying the legal responsibility boundaries of data providers, developers, and users. From the perspective of responsibility allocation, legal mechanisms should be established to force relevant parties to be cautious in the process of data construction and use, preventing situations where output results are unreasonable, illegal, or unsafe due to data defects, and avoiding the consequences of data defects. Different artificial intelligence products should be subject to national and industry standards for data acquisition, use, and classification and grading, with usage levels ranging from unrestricted, restricted, permissive, to prohibited. A specialized agency should be established to supervise the artificial intelligence industry.

Key words: artificial intelligence; data defect; legal risk prevention; distribution of responsibility; supervision and administration

(责任编辑:陈思婷)