收稿日期:2023 - 12 - 28

生成式人工智能的法律风险及侵权责任认定

严雅茹,骆小春
(南京工业大学 法政学院,江苏 南京 211816)

摘要:生成式人工智能迅猛发展的背后折射出科技进步与法律制度发展的辩证关系。当前生成式人工智能技术带来众多科技红利的同时,也带来诸多法律风险,主要存在于个人信息保护、生成内容的版权问题等方面,并由此带来侵权责任制度的选择问题。秉承鼓励创作与创作者权益保护的理念,应当认可一定情形下生成式人工智能生成内容具有可版权性,并为版权权益归属、版权侵权情形下侵权主体的界定提供依据。综合考量各主体的风险控制能力与安全防范义务,生成式人工智能的侵权责任应当采取一般侵权责任制度并辅以过错推定规则。基于生成式人工智能侵权的特殊性,应当对服务提供者变通适用“通知-删除”规则。

关键词:生成式人工智能;可版权性;产品责任;过错责任原则;过错推定
中图分类号:D923.1　　**文献标识码**:A　　**文章编号**:1003 - 6873(2024)02 - 0054 - 12

作者简介:严雅茹(1999 —),女,安徽合肥人,南京工业大学法政学院硕士研究生,主要从事民商法学研究;骆小春(1969 —),男,湖北黄冈人,南京工业大学法政学院教授,博士,硕士生导师,主要从事民商法学研究。

DOI:10.16401/j.cnki.ysxb.1003 - 6873.2024.02.020

人工智能技术是依托互联网蓬勃兴起的数字技术的产物,人工智能可细化为分析式人工智能与生成式人工智能。前者主要是从大量数据中挖掘隐藏模型并进行一定预测,后者则能够基于大数据学习,凭借算法模型技术自主生成文本、图片、视频等新内容[1]。二者的区分在于后者属于单纯的人工智能系统,基于强大的数据库自主学习并发现语言规律,在与用户的互动中生成“内容”或创作“作品”。2022 年 11 月,美国 Open AI 公司研发的 Chat GPT 进入公众视线,标志着生成式人工智能技术的发展取得里程碑式的突破。Chat GPT 集智能搜索引擎、智能文本分析器、洗稿器于一体,其本质上为大语言模型。为确保人工智能系统的安全性,2023 年 7 月,国家互联网信息办公室等相关部门发布《生成式人工智能服务管理暂行办法》(以下简称《暂行办法》)。作为我国首部对生成式人工智能进行治理的部门规章,其规定了人工智能服务提供者的多项义务和服务方面的要求,坚持了安全与发展并重、审慎包容与分级分类监管并行的思路。生成式人工智能基于其技术复杂性、系统高度自主性以及生成内容不确定性等复合属性导致了其在适用过程中仍会不可避免地带来损害,引发法律与伦理的双重风险。当前生成式人工智能发

展的关键点在于如何把控安全发展与科技发展之间的方向盘,在秉持安全发展理念的同时,营造科技向善、创新优化的市场大环境。基于该领域的复杂性和相关规范的缺位,本文从生成式人工智能发展面临的法律风险入手,如在数据的收集与处理过程中可能会侵犯个人隐私、生成内容属于虚假有害信息、生成内容侵犯他人知识产权等,并在此基础上进一步讨论生成式人工智能涉及侵权时责任的认定问题。

一、生成式人工智能面临的法律风险

(一)生成式人工智能引发个人信息安全风险

1. 生成式人工智能收集利用个人信息带来的挑战

生成式人工智能需要以大量的数据为支撑进行模型训练,数据的合法获取与使用是生成式人工智能面临的首要风险。具体而言,主要体现在以下几个方面:一是个人数据的合法性获取,在数据的收集处理环节获取数据未经信息主体许可。大多数国家(地区)均承认个人信息的收集应当遵循“知情同意规则”,我国也不例外。相关主体收集个人信息时应当通过其隐私政策向用户告知收集其个人信息的行为,由信息主体决定是否同意此种信息收集行为。根据 Open AI 官方网站公布的隐私政策可以看出,该公司在个人信息的收集过程中并未严格秉承知情同意规则[2]。此外,由于生成式人工智能系统训练模型的需要,可能涉及第三方数据源的访问。生成式人工智能相较于先前的人工智能的进步之处在于其高度自主性,可以在脱离人类监督的情形下主动抓取互联网数据并进行筛选,并对其收集来的大量数据进行关联分析与处理。依据《中华人民共和国民法典》(以下简称《民法典》)第一千零三十二条的规定,隐私的判断标准不再以内容公开与否为依据,权利主体不愿为他人所知晓的公开私密信息仍属于“隐私”所涵摄的范围内,侧重于强调权利主体的主观意愿。退而言之,即使其收集的为非私密信息,生成式人工智能也可在运行阶段自行将该信息整合,增强信息主体的可识别性,威胁社会不特定公众的隐私安全。二是个人数据非法使用风险。生成式人工智能在利用个人信息的过程中,难以严格遵守《中华人民共和国个人信息保护法》(以下简称《个人信息保护法》)中规定的目的限制原则与公开透明原则,《个人信息保护法》第六条明确信息处理者在处理个人信息时不得违反先前收集个人信息时所具有的“明确、合理的目的”,不得超越信息收集时所主张的处理目的。而 Open AI 公司的隐私政策则明确指出,在某些情形下其会将个人信息提供给第三方,比如在用户不知情的情况下将用户带有消费倾向的个人信息提供给第三方商业利用。另外,《个人信息保护法》第七条和第二十四条所规定的公开透明原则则要求信息处理者对外披露个人信息处理的细节。由于生成式人工智能存在“算法黑箱”,基于数据所生成的内容存在着包括设计者在内的主体都不可洞悉的属性,更无从谈及对外披露个人信息的处理细节,由此便造成对公开透明原则的违背。即生成式人工智能在数据收集环节后的使用环节尚存非法泄露个人数据的风险,此种大范围传播公民个人信息甚至还会引发国家数据安全。目前,Chat GPT 平台已经出现对话数据与相关信息的泄露情况,2023 年 3 月意大利数据保护局也因此对外宣称封锁 Chat GPT。

2. 生成式人工智能生成内容引发虚假有害信息风险

自 2011 年起,意大利的一名商人以谷歌搜索引擎的自动补足功能侵害其名誉权为由向谷歌发起诉讼。此后,由算法所引起的虚假信息诉讼已屡见不鲜,但争议双方都围绕虚假信息是否由算法生成而展开攻防,侵权要件的构成都较为单一。在人工智能尚未普及的时代,虚假信息便会

因为其极具迷惑性而造成社会的混乱,生成式人工智能以大型数据为支撑所生成的虚假内容更难以辨别[3]。《暂行办法》作为我国首个AIGC监管文件,其第四条明确规定:不得生成虚假有害信息,侵害他人的名誉权、荣誉权等,强化了监管生成式人工智能抵御网络虚假信息的重要性。由于生成式人工智能所引发的虚假有害信息风险的内在机理和表现形式与自动补足算法有所不同,所以需要具体分析。

从致害的原因上看,生成式人工智能的虚假信息侵权主要有以下两个方面:一是由其自身模型导致。由于Chat GPT并不提供多来源的信息参考,且其无法实现真正意义上的推理,在生成式人工智能系统被问及的问题在其数据库中欠缺相关资料时,由于其系统的高度自主性,其通常会编造虚假信息或嫁接其他内容,可能会生成看似合理实则虚假或误导性的信息。生成式人工智能系统算法在训练过程中可能将人类作品中的隐性或显性偏见予以继承,从而导致生成内容存在价值偏见、与当下主流价值观相悖甚至生成内容具有歧视性、侮辱性色彩。人工智能逻辑包装下的信息难以被人类辨别真伪,基于庞大的用户群体,一些细微的疏漏就会使得该虚假信息被用户广泛误用和传播。二是由使用者操作不当导致。人工智能具有弱可控性,生成式人工智能的提供者也无法预料算法模型将会产生何种结果,从而会造成使用者的盲目信任甚至产生人机合谋的风险。恶意使用者借助生成式人工智能编造虚假新闻、文章,误导公众,如2023年5月平凉市公安局网安大队侦破的我国首例利用生成式人工智能编造虚假信息的案件,犯罪嫌疑人通过恶意炮制Chat GPT收集到的新闻要素并上传至自媒体平台,引起舆论广泛关注。此外,生成式人工智能系统经过人类的反馈,不断调整输出结果,后续的输出内容则会受到前次用户输出内容的影响,系统的此种特性也给恶意的谣言制作者留下了操作空间。

不同原因导致的生成式人工智能生成内容引发的虚假有害信息风险问题早已引发各界关注。欧洲议会在《人工智能法案》的修改过程中更是将“缺乏真实性”规定为生成式人工智能的核心风险[4];我国《暂行办法》第四条所规定的“不得生成虚假有害信息”亦是体现了对生成式人工智能生成内容引发虚假信息的重视。由于生成式人工智能生成内容侵权风险可预测性较低,故在损害实际发生的情形下应当考虑侵权情形下法律制度的选择,按照生成式人工智能的法律定位使相关主体承担侵权责任,在遵循法治思维保持谦抑性的同时,实现生成式人工智能技术的稳步发展。

(二)生成式人工智能生成内容的可版权性之争

当前的数据环境对现行知识产权制度构成巨大挑战,人工智能生成内容的版权问题面临新的法律挑战。目前,生成式人工智能主要分为三种类型:一是由人工智能独立创作;二是由人类使用者辅助创作;三是按照人类使用者输入的内容生成[3]。学术界对于生成式人工智能生成内容是否具有可版权性存在明显的理论分歧。支持者以“人工智能为作品创作工具说”为理论基础,认为生成式人工智能系统通过与人类的互动融合了人类的主观倾向。生成内容在一定程度上体现了人类智慧,只要满足《中华人民共和国著作权法》(以下简称《著作权法》)中关于作品的要求且不涉及特殊情形时,应当将该生成内容视为著作权法意义上的作品[5]。反对者则认为,以Chat GPT为代表的生成式人工智能所生成的内容不符合著作权法的保护对象,现行著作权法主要保护人的智力成果[6]。另外一种折中观点则认为对此不可一概而论,应当综合考察该生成内容是否在一定程度上体现创作者的精神活动,该精神活动能否被评价为著作权法上的创作行为[7]。

人工智能生成内容可版权性的界定,不仅关乎生成内容的版权权益归属,更关乎出现著作权侵权时责任分配问题。对人工智能生成内容可版权性的界定不明,将直接影响该生成内容的版

权权益归属:若认可人工智能生成内容具有可版权性,那么该作品就应当归属该人类创作者;若认为该生成内容不具有可版权性,则需要在生成式人工智能提供者与人类创作者之间构建合理的权益分配规则。此外,对于人工智能可版权性的界定不明将导致该生成内容侵权情形下,侵权责任的分配不明确。基于生成式人工智能技术高速发展,其用于训练的海量数据不断更新,在此过程中自动抓取和解析他人作品,使生成式人工智能的生成内容可能涉及与他人作品相类似甚至抄袭他人作品的行为,或者未经许可不自觉地改编了他人的原创作品,极大可能会诱发对先前作品的版权侵权风险。在通常情况下,行为主体应当对其所实施的侵权行为承担侵权责任。经由服务使用者进行操作导致生成内容侵权的情形下,由该使用者承担侵权责任具有合理性,但由于生成式人工智能的提供者存在算法设计缺陷而使得生成内容侵权时,该提供者也可能据此承担一定责任。此时,该侵权责任应当如何分配呢?

综上所述,对于生成式人工智能生成的内容是否具有可版权性、版权权益的归属以及在该人工智能生成内容构成侵权情形时,该侵权责任应当由服务提供者还是使用者承担,在学界存在较大争议。鉴于相关规范的缺位,所以亟需从理论上对该问题予以回应。

二、生成式人工智能引发个人信息安全风险之预防

相较于普通个人信息侵权案件,因生成式人工智能所引发的个人信息安全风险具有颠覆性与不可挽回性,所以对于生成式人工智能引发个人信息安全风险的规制不能仅沿用传统路径将关注点置于损害发生后的救济,而应侧重于风险预防。欧盟的《数字市场法》堪称国际社会数字平台事先治理的立法典范,其强调以事先预防代替事后追责,从源头上遏制因数字治理失衡所带来的负面效应,对我国亦有不可忽视的指导作用。《中华人民共和国数据安全法》第二十二条明确要求建立“统一集中、高效权威的数据安全风险评估,信息共享、监测预警机制”。作为新态势下出现的新型生产要素,数据是我国数字化、网络化、智能化的基础[8],在当前的社会背景下数据合规被赋予新的内涵,对于防范数据安全风险具有无与伦比的功能。数据合规机制的构建应当从客体保护与多元化治理两个方面予以考量。

(一)加强个人数据保护

1. 个人信息的匿名化处理

在生成式人工智能收集信息的过程中,一旦泄露个人信息,将使信息主体面临巨大挑战。通常认为,个人信息的核心要素在于信息的可识别性。依据《中华人民共和国信息保护法》第七十三条的规定,去识别化主要有“去标识化”与“匿名化”两种手段。二者的区分在于:经过去标识化手段处理的个人信息只是被识别出来的可能性降低,其与其他信息经过整合依旧能够识别出信息主体;而经匿名化处理的个人信息则较为彻底,具有不可逆转性。故生成式人工智能的提供者应当采用匿名化手段。在具体措施上,生成式人工智能服务提供者可以通过采用假名、加密、哈希函数等技术手段将个人信息中的直接标识符或准标识符删除或变换,去除标识符与信息主体之间的关联性成果[2]。诚然,从理论上说,凭借匿名化信息可能也能复原出个人信息甚至个人敏感信息,但在需要花费高昂的成本与时间进行复原且与之所欲达到的目的不成比例时,会极大地阻碍人们对匿名化信息的分析行为。

2. 保障数据质量

我国《暂行办法》第七条将“保障数据质量”作为生成式人工智能风险防范的重要原则,力求

增强训练数据的客观性、真实性,进而保障原始数据与生成内容的质量。信息收集主体对于原始数据应当确保其收集过程合法,通过算法设计对所收集来的数据进行准确标示,明确数据的来源、类别等本质属性。评估数据的风险等级,采取更加严格的保密措施与存储介质。对于其生成内容而言,应当建立数据安全评估机制。优化算法与模型,综合考量数据来源、事实审查等,验证生成内容是否具有真实性与可信度,降低虚假信息风险,以提高生成式人工智能的数据质量。

(二)构建多元化数据合规机制

数据合规管理应当坚持事先防范与事后救济两个方面。首要问题是应当如何防范风险,构建风险预防规则阻碍风险的现实发生。无论是互联网技术公司还是利用人工智能进行数据收集的互联网企业,都应对其数据信息进行数据安全合规管理以达到数据风险防范的效果。数据风险的事先防范方面,应当协调建立数据合规的多元化治理格局。一是完备企业数据合规文件。为公司的数据治理提供规范层面的依据,在算法程序中标注规范层面的要求与义务。二是健全配套适用规则。对信息主体的个人信息收集行为予以通知;个人信息的利用环节应当取得用户的明示同意,由此限制生成式人工智能服务提供者对该数据的使用目的;将信息利用的主要环节向信息主体披露并告知该信息处理行为可能存在的风险。三是建立专业性数据合规团队。吸纳精通法律、算法等交叉学科背景的复合型人才。一方面督导企业数据合规事务按照流程进行,增强数据的隐私程度;另一方面实现企业内部职能部门、行政机关以及第三方机构的协作,共同构建涵摄多主体的多元治理格局。四是建立数据监管体系,建立严格的准入门槛与备案制度。对数据的训练和生成全流程进行控制,对使用者的使用行为进行检测,防止数据的滥用。

对于数据风险的事后救济,企业应当对因数据泄露所带来的不利后果进行类型化分析,根据法律行为与法律责任承担的后果界定行政、民事与刑事责任。通过行政处罚、多元化民商事纠纷解决机制、刑事追责等手段,及时进行合规整改,降低数据安全风险带来的损失,最大程度地降低负面影响。

三、生成式人工智能生成内容的可版权性认定规则

正如上文所列,关于生成式人工智能生成内容能否被认定为具有版权性,各学者莫衷一是。随着生成式人工智能内容与人类创作者主观意识的高度融合,认定生成式人工智能作品满足特定情形时具有可版权性,不仅契合了《著作权法》第一条所蕴含的鼓励作品创作与传播的理念,而且有助于推动科技创新发展、激发文化领域的创作积极性,为社会主义的繁荣发展助益。

(一)生成式人工智能生成内容的可版权性认定

1. 生成内容的独创性界定

“创作”是构成“作品”的充分必要条件,创造性则被界定为判断作品是否具有独创性的核心要素。早在1903年美国法院经过判例首次明确独创性的定义:认为若该生成内容经由作者独立完成且属于版权法所规定的种类,那么该生成内容则具有独创性[①]。即作者只需满足最低限度的创造力,对于独创性的要求较低。2023年3月,美国版权局亦是率先对生成内容的可版权性问题正式做出回应:认为以Chat GPT为代表的生成式人工智能所自动生成的内容不受版权法的保护,即便人类使用者在此过程中存在以指令输入等方式的行为,但其训练的数据仍是基于人

① See Bleistein v. Donaldson Lithographing Co. 188 U.S. 239, 23 S. Ct. 298, 47 L. Ed. 460.

类的创作作品,所以不具有创作价值。只有该生成内容包含足够的人类创作者的创作要素,才可受到版权法的保护。我国《著作权法》亦是明确作品构成的四要件①,其中“独创性”属于作品认定的核心与关键,即判断生成式人工智能作品是否受到著作权法保护应当以著作权法中的“独创性”为出发点,综合判断以 Chat GPT 为代表的生成式人工智能作品是否具有独创性、是否属于人的“智力成果”。目前看来,在对生成式人工智能作品进行著作权法意义上的作品评价时,独创性的认定主要考量以下方面[9]:一是内容是否具有创新。我国著作权法保护的并非思想,而是作者对于内容的表达。即便是基于相同理论学说,但表达者在表达上方式上别具一格,凸显了新颖性和独立性,其仍可界定为具有独创性。二是独创性对于作品的艺术、文学价值等并无程度上的高低要求。三是独创性只需要体现“一定程度的创造性”,即该作品不是对现有作品的简单复制或抄袭即可。四是独创性的认定实则是与现有作品的相对比较。即不要求该作品的产生是从无到有,如果该作品对于现有作品在表达上具有一定程度的差异性,那么该作品也具有独创性。目前的生成式人工智能的运行机理在于:其算法被设计出来后,辅以大量的数据对其进行训练,从而不断地生成内容。这就意味着部分生成式人工智能作品虽然在内容上有相同或相似之处,但它们在数据整合过程中抽离出些许先前尚未察觉的规律与趋势而赋予了作品新的内涵,体现了一定程度的创造性,所以该生成内容也可以认为具有独创性。此外,即使其进入使用阶段,还是需要不断供给新的数据以满足其可能面对的新场景的需求,则可能会涉及对其他作品的引用与整合,对于此种特殊类型作品如数据库、编译作品等,独创性的判断则需综合考量作品引用与被引用部分是否符合著作权法的比例要求、该作品是否对其引用的作品进行了一定程度的改编与创新、是否体现了作者的智力劳动与智慧等。

2. 生成内容的主体资格

在通常情形下,生成式人工智能所生成的内容符合著作权法中客观上的“独创性”要件,但在著作权的主体资格方面仍存在法律障碍。在 1956 年前后,美国版权局与联邦法院先后认为作品的创作者必须是人类。在 2018 年 David Slater 案中,动物组织主张猴子应当对其拍的照片享有著作权。美国第九巡回法院通过对《版权法》《美国版权局实践纲要》等有关规定进行解读,最终未予承认动物的版权主体资格[10],该案判决对非人类作者是否可以成为版权主体这一问题具有重要指导意义。申言之,美国在著作权主体资格的界定方面,前后并无实质性变化,均认为作品的创作者必须是人类。我国司法实践中对于人工智能生成内容的可版权性则采取不同的保护路径,在肯定涉案生成内容具有独创性的基础上对于主体资格的认定却大相径庭。北京互联网法院以人工智能不属于法律主体而否认了人工智能生成内容享有著作权②。生成式人工智能既不属于“人”,也不具有类似于法人的拟制人格的属性,所以不能成为著作权法意义上的主体,其只是经凭人类创造出来的一项科技成果[11]。无独有偶,在我国首例认定人工智能生成内容构成作品的生效案件中,深圳市南山区人民法院则在独创性判断与创作过程的分析方面做出重要探索,在认定生成文章具有独创性的基础上从涉案文章的生成过程来分析是否体现了软件主创团队个性化选择等因素,并据此认可了人工智能生成内容的可版权性③。区别于受《著作权法》保护的传统作品的认定,评价生成式人工智能作品是否构成《著作权法》意义上的作品时,需要考量人类

① 《中华人民共和国著作权法》第三条明确规定了作品构成的四要件,即作品为“文学、艺术和科学领域内”具有“独创性”,“能以一定形式表现”的“智力成果”。

② 参见北京互联网法院民事判决书(2018)京 0491 民初 239 号。

③ 参见广东省深圳市南山区人民法院民事判决书(2019)粤 0305 民初 14010 号。

创作者是否对生成式人工智能作品的独创性起到突出的实质性作用[12]。生成式人工智能系统未经人类贡献而自动生成的内容不属于“作品”,比如法律、法规以及对相关事实的无独创性汇编等[9],不受著作权法的保护。诚然,在作品的创造过程中,生成式人工智能能够在强大的数据库中挖掘可用信息,在理解人类创造者输入内容基础之上,创作出与之相关的、有意义的新内容。这种“输入-‘理解-转化’-输出”的创作方式[11],为人类创作者提供了灵感,对拓宽著作领域、突破思维定式有所裨益,极大地丰富了作品的内涵与形式,为人类创作者提供了不可估量的支持,但仍无法改变其所表征的“工具性”。人类创作者经过筛选挑出最符合自己构思的创作意图,美化作品使其符合自己的审美,使该内容具有著作权法意义上的独创性,在整个创作过程中人类创作者仍占据核心地位。

综上所述,对于生成式人工智能生成内容的可版权性认定不可一概而论。在人类使用者直接参与的情形下,当该生成内容满足著作权法意义上的独创性要件并体现出人类创作者在此过程中的实质性贡献,其才有可能成为著作权法保护的对象。在目前的法律框架下,认可生成式人工智能生成内容的可版权性,则可为生成式人工智能生成内容的权益归属与侵权责任主体的认定提供依据。

(二)生成式人工智能生成内容的权益归属

在一般情况下,依照经济学上“谁投资,谁收益”的原理,人类创造者对于该生成式人工智能生成的内容具有实质性贡献,应当享受由此所带来的收益,故该版权应当由该人类创造者即生成式人工智能系统的使用者所有。生成式人工智能系统在此过程中仅充当该使用者的创作工具,若该系统的开发者与提供者享有该版权归属,则违反了《著作权法》中作者权益保护的理念,失衡的权益分配机制更是对积极的创作氛围的扼杀,使得创作者的合法权益无法得到保障。Open AI的使用条款亦是明确指出,该产品的输出内容在用户遵守条款的前提下,由Open AI将所有权利与利益转让给用户。也有学者认为按照投入与产出关系应当成正比的理念,生成式人工智能系统的开发者在设计程序与优化系统的过程中所进行的调整训练数据、优化算法模型等行为对于该生成内容的可版权性也可认定为存在一定贡献。对于作品的版权权益的分配,在创作者与开发者协商一致的情形下,应当允许双方约定按照各自贡献进行版权权益分配[9]。此种看法具有一定的合理性。依据我国《著作权法》第十四条的规定,合作作者有权自主协商该作品的权益分配,此种合作关系的建立更是有利于促进技术的进步与版权合作意识,是各方合作共赢的体现。借鉴合作作品的权益分享规则,认可双方协商约定共享该版权权益的情形,不仅可以补偿开发者的开发成本,激发产业活力,而且可以分担版权侵权情形下所应当承担的责任。

四、生成式人工智能侵权责任界定之分析框架

法学的使命不在于赞赏因科技发展所带来的辉煌成就,而是应当检视科技发展可能导致的非理性后果,进而通过法治降低科技发展的风险[13]。面对生成式人工智能此种新技术所引发的法律风险应当如何应对呢?首先,民法秉承鼓励创新、风险预防的价值取向,此种价值取向细化到侵权责任编内,更侧重于预防损害的实际发生,而并非一味制裁。目前大多数国家对于生成式人工智能的讨论主要侧重于风险预防的角度,如欧盟的《人工智能法案》以及我国的《暂行办法》。通过将潜在风险控制在可接受范围内的事前治理模式虽然在一定程度上能为生成式人工智能的规范发展赋能,但生成式人工智能承载的高科技使其所带来的风险具有不可预测性和不确定性。生成式人工智能的侵权类型主要包括:第一,在信息收集过程中,生成式人工智能服务提供者未经信息主体同意,收集涉及个人隐私的数据用以训练,此种信息收集与处理行为可能构成侵权;

在信息使用过程中,由于算法原因或人类使用者恶意使用生成式人工智能系统,致使其生成虚假有害信息,对其进行传播可能构成侵权。第二,生成式人工智能的生成内容可能存在抄袭、改编他人作品的行为,可能会造成对他人作品的版权侵权。在侵权损害后果实际发生后,所需要解决的问题则为损害的分配,此时便涉及侵权责任的界定。

(一)特殊侵权责任分析路径

立足于生成式人工智能领域,生成式人工智能之所以能从其他法律客体中脱颖而出得到学术界的广泛关注,得益于其算法设计,生成式人工智能侵权也大多是由于其算法设计缺陷所致。通常认为,产品缺陷包括生产缺陷、设计缺陷和经营缺陷。我国《民法典》规定,“因产品存在缺陷造成他人损害”的适用无过错责任原则,并未对缺陷情形的归责原则予以细化。在当前的法律背景下,学界对于能否通过民法中的产品责任来防控和救济生成式人工智能技术发展所带来的风险,尚未达成一致意见。诚然,生成式人工智能的规范发展似乎更需要严格责任来保驾护航,以此最大程度保障受害者的权益。但算法设计作为尖端科技领域的领头羊,在实际发生损害之前,设计者只能在现有认知范围内予以设计,其潜在风险难以被设计者所识别。在我国产品责任的制度中,在产品投入流通时,因当时的科技水平不能发现该产品存在缺陷,行为人有权据此免责。由此所带来的现象是,在生成式人工智能算法设计缺陷的追责中,大部分设计缺陷将得以免责,极易造成该现有技术抗辩的滥用,无过错责任归责原则的确定难以达到规范意旨。

从目前的产品责任制度出发,我国《民法典》规定产品责任制度的适用对象为“产品”。由此所带来的首要问题为生成式人工智能系统是否属于产品责任法上的产品?我国《民法典》以及相关产品质量法对此并无明确规定。对于此,美国则将软件区分为产品与服务两类,若该软件可经由大规模的复制与销售,具有高度的同质性,则被认定为产品。反之,若该软件具有针对性,属于为客户量身定制的,则属于服务。产品责任属于特殊侵权责任,适用无过错责任原则,与一般侵权责任的基本特征不同,产品责任不以产品制造者的过错为要件。产品制造者之所以需要肩负起严格责任,是因为其可以凭借市场的价格机制来分散其所应当承受的赔偿责任,生产者将需要承担的责任加入产品中,向市场大量销售同质化产品,由市场消费者分摊产品所造成的损害。对于生成式人工智能而言,其所生成的内容受到人机交互的差异与算法模型变化的影响,致使其个性化服务特征凸显,欠缺高度的大众化特征,不具有普适性,所以难以将其界定为产品。申言之,从文义范围来看,生成式人工智能系统无法适用产品责任。

此外,产品责任的另一特性在于其损害赔偿责任的承担范围突破了合同的相对性。生成式人工智能系统的应用范围涉及生活的方方面面,用途广泛。若对其科以产品责任,那么生成式人工智能系统的提供者一方面需要对与其有合同关系的用户承担损害赔偿责任,若存在涉及第三人的情形,即人工智能系统的使用者以该系统为自己的客户提供专业服务,如律师、会计师等,生成式人工智能系统的提供者也需要对合同以外的第三人承担责任,那么生成式人工智能系统的提供者所承担的风险是难以想象的。目前生成式人工智能技术处于发展阶段,若采取类似欧盟的《人工智能法案》对人工智能提出多项强制性要求的路径或对生成式人工智能系统的提供者科以过于严格的责任,会使人工智能的提供者为规避自身责任风险而对系统的迭代升级进行限制,这样必然会阻碍生成式人工智能系统技术的创新与进步。

在规范层面,国家互联网信息办公室在《暂行办法》(征求意见稿)中,存在“生成式人工智能产品”“生成式人工智能服务”等区分,而在正式颁布的《暂行办法》中仅见“生成式人工智能服务”的表述,将生成式人工智能定性为“服务”。此举也可说明在我国目前尚无为生成式人工智能系统设立特殊侵权责任的立法趋向,对于侵权责任路径的分析也具有一定的启发意义。

(二)一般侵权责任分析路径

对于生成式人工智能致人损害侵权责任的界定,除考虑产品责任此种特殊侵权责任外,还可以以一般侵权责任路径为出发点。例如,美国将缺陷类型予以分类,并适用不同的归责原则。美国起初对于所有的产品缺陷均采严格责任,后经司法实践需要,先后通过1979年《统一美国产品责任示范法》、1997年《侵权法重述第三版:产品责任》正式确立设计缺陷的过错责任原则,明确认定标准为判断行为人是否存在过错的合理性标准[14];欧盟委员会关于人工智能侵权责任的立法建议则包含以过错责任为基础的一般侵权责任的归责路径[15]。法律秉承保护相关主体生命、健康等权利而对产品损害科以严格责任,但考察因生成式人工智能发生损害的主要情形,生成式人工智能通常不会直接造成人身或财产等物质性损害,损害通常出现在侵害名誉权、著作权等领域。在提及人工智能算法侵权责任时,主观过错无不作为重要考量因素。归责的基础是自由意志,人工智能构成了对自由意志的极大挑战,同样对于法律责任的认定产生极大影响[16]。基于生成式人工智能侵权原因的多样性,应依据具体情形划分各主体的责任承担,谨慎使得各方主体承担连带责任。因而以在一般侵权责任框架下的过错责任原则为基础具有合理性。

我国《民法典》规定一般侵权责任以过错为归责原则。若生成式人工智能适用过错责任原则,应当需要满足该侵权责任构成要件。只有当生成式人工智能服务提供者存在过错时才承担损害赔偿责任,且一般由原告举证证明被告存在过错。通常认为,过错包括故意与过失两种主观心理态度,故意又可划分为直接故意(明知+希望)与间接故意(明知+放任)。生成式人工智能生成内容具有随机性与不可预见性,所以构成直接故意的主观心理态度的可能性较小。比如被侵权人发现系统所生成的内容侵害其著作权,并向生成式人工智能系统的提供者发出通知,那么则可以推定提供者对于损害是明知的。有观点认为,生成式人工智能服务提供者的过错应当根据不同情形予以分别认定:即以"一般理性人"为标准,若服务提供者尽到了一般理性人的注意义务则其无需承担责任;若人工智能产品的生产者、销售者以及使用者均存在过错,则三者承担按份责任或连带责任。上述观点看似具有合理性,实则不然。一方面,对于"一般理性人标准"的认定,由于生成式人工智能系统随着与用户间的交流互动日益频繁、数据库的补充,其模型参数不断更新,算法结构日趋复杂,且通常承载高科技性质,科技壁垒较强。若其生成内容含有侵犯他人权益的内容,由使用该系统的一般公众举证证明该生成式人工智能系统存在设计上的缺陷或证明生成式人工智能系统的提供者对损害结果的发生存在过失,显然难度较大。申言之,在因个人信息侵权的情形下,信息主体对于生成式人工智能系统的认知程度难以支持其提供证据证明生成式人工智能服务提供者及使用者的侵权行为。相较于此,此时采取过错推定的方式,将举证责任转移至被告更具合理性。由生成式人工智能服务提供者与使用者就其不存在侵权行为予以证明,生成式人工智能服务提供者举证证明其违反义务的行为与损害结果的发生不存在法律上的因果关系,便可以推翻上述规定[4]。退而言之,即使采取过错推定的归责原则,该规定也是可以被推翻的[15]。另一方面,此种观点实则是将生成式人工智能适用产品责任而认定各责任主体的责任,忽视了生成式人工智能产品与产品责任上的产品不可相混淆的重要前提。部分学者基于传统的"理性人"标准提出"理性计算机"的标准,认为应当将人工智能类比于人,依据计算机系统的行为来判断其是否存在过失,并在此基础上认为"理性人标准"与"理性计算机标准"最终将融合为一体[17]。不可否认,上述观点有一定的启发性,但其不合理性显而易见。以Chat GPT为代表的生成式人工智能为例,尽管其在大多数任务上展现出超越人类的水平,但也在某些情况下出现在人类看来极其低端的错误,比如不准确的信息、误导性的言论,以及不符合价值观的歧视性话语等,所以难以用一个相对固定的标准来衡量人工智能系统是否尽到合理的注意义务。生

成式人工智能算法运行的高度自主性对传统的“主体-行为-责任”追责机制提出了挑战,生成式人工智能作为由人类设计与开发而形成的工具或技术,其并不属于拥有权利和责任的实体。而我国《民法典》所规定的主体资格,则是指能够独立的行使权利、承担责任、得到法律认可的实体,换言之,其既不属于“自然人”也不属于法律上的“拟制人”。由于生成式人工智能不具备法律主体资格,也无法对其行为承担责任,所以此处的过失实则是人工智能提供者的过失。当人工智能致人损害时,仍需综合考量生成式人工智能的服务提供者在该系统的设计、研发、检测过程中是否尽到了合理的注意义务,是否满足《暂行办法》对生成式人工智能服务提供者所规定的各项义务和要求。当该系统的提供者未严格遵守上述规定,那么就可推定其对损害结果的发生存在过错的主观心理态度。

基于生成式人工智能系统与用户之间的交互性,用户的不当引诱行为也可能是导致生成式人工智能系统产生侵权内容的因素之一。在用户的不正当引诱使得系统所生成的内容构成对第三人的侵权时,用户的诱导行为虽存在一定的过错,应当受到法律的谴责,但并不必然导致提供者就此免责。即第三人的过错不能阻断系统提供者的行为与损害结果之间的因果关系。对于系统提供者而言,应当考量其对用户的不正当诱导行为是否尽到了安全保障义务,在监测到风险时是否及时处置。在未尽到风险防范的安全保障义务时,系统提供者应当就其过错与该用户对第三人共同承担侵权责任,二者构成共同侵权行为。在被侵权人向系统的提供者主张侵权损害赔偿时,提供者有权主张受害人对损害的发生存在一定过错,从而减轻或免除自身应当承担的赔偿责任。此外,《暂行办法》第九条规定的“服务的提供者承担产品生成内容生产者的责任”,并未考虑特殊情形下的责任归属。在生成式人工智能生成内容造成版权侵权情形下,若认可生成式人工智能生成内容具有可版权性,那么依据现行法律框架中版权的归属规则,则能够为侵权责任的归属提供正当性依据,确定版权侵权情形下侵权责任的承担主体。具言之,在一般情况下,人类创作者单独享有该生成内容的版权权益,并就该版权权益可能出现的侵权责任承担不利后果;在特殊情形下,生成式人工智能系统开发者与人类创作者基于双方协商一致而按照比例共有该版权权益,那么双方就应当在各自享有版权权益的范围内向被侵权人承担侵权责任。此种特殊分配方式有利于平衡各利益主体的权益,促进生成式人工智能在各领域的蓬勃发展。

由此带来的另一问题为,生成式人工智能系统是否适用现行法律框架下关于网络服务提供者责任限制的规定。早在互联网兴起之时美国通过制定《通信净化法》第二百三十条为网络服务提供者做出免责规定,并在其后通过《数字千禧年版权法》确立“通知-删除”规则,在版权法领域为网络服务提供者提供“避风港”式保护,上述规则的设立对欧盟、中国等相关国家的网络立法起到了极大的促进作用。我国《民法典》亦是对传统网络侵权采取的“通知-删除”规则予以规定,并详细区分了网络服务提供者因自己实施的侵权行为、因其用户实施的侵权行为所需承担的责任。考察因搜索引擎的补足算法侵犯名誉权案例的主要情形,责任承担标准主要为技术中立原则免责、严格责任和过错责任三种。目前我国对“技术中立原则”有“实质性非侵权用途”与“技术工具论”两种理解[18]。实质性非侵权用途是指一个产品如果可以被广泛用于合法用途,那么该产品的提供者则免于承担共同侵权责任。其本质上是一种法律解释原则,用以解释权利人与传播人用来分配新型传播技术而产生的利益,而并非用来强调技术的提供者因为其技术具有实质性用途而免于承担侵权责任;第二种是技术工具论,即采用“技术本身中立”但“技术应用不中立”的二元归责体系,我国理论界与实务界多采此种说法①。比如,搜索网络服务的提供者所提供的推荐

① 参见北京知识产权法院(2020)京 73 民终 1914 号民事判决书,浙江省杭州市中级人民法院(2018)浙 01 民终 7312 号民事判决书。

词属于网络用户在特定时间的搜索词的动态反映，搜索服务的提供者对该搜索结果仅为客观和中立的平台服务，不存在“主观过错”，因此不承担侵权责任[19]。但生成式人工智能系统所生成的侵害他人权益的内容并非该系统的使用者通过系统自行创作和发布，相反，将该侵权内容看作是系统提供者创作与发布的信息可能更具说服力。在此情形下，如果将生成式人工智能系统的提供者予以免责或限制责任，那么受害人的损失将无从救济。另外，与此处有所不同的是，由于生成式人工智能生成机制的不可预测性与生成内容的随机性，其无法直接采用删除、屏蔽和修改输出内容来消除影响。生成式人工智能系统服务的提供者只能及时采取有效措施消除侵权信息所带来的影响并再次优化系统，进而实现算法的更迭换代，防止再次生成侵权信息。综上所述，在生成式人工智能领域，现行的“通知-删除”规则应当予以变通适用，其性质应当属于义务与责任的构成条款，而并非生成式人工智能服务提供者的免责条款。

以 Chat GPT 为代表的生成式人工智能技术的创新发展方兴未艾。当前社会对于生成式人工智能的期待为在保留其积极社会效益的基础上尽量降低其带来的负面效益，为保护个体权益而为科技发展套上枷锁似乎并不妥当。法治是治国的重器。作为上层建筑范畴，法律制度应当能动地对技术的进步和创新做出回应，为生成式人工智能在中国特色社会主义法治框架下规范健康发展保驾护航。同时，应当遵循法治运行的基本原理与科技创新的发展理念，寻求各主体间的平衡。在生成式人工智能生成内容的可版权性认定上，应当综合考虑该生成内容的独创性以及人类创作者的实质性贡献，并在此基础上明确该版权权益归属与侵权主体；在因生成式人工智能侵权所带来法律政策的选择上，应当以一般侵权责任制度为基础，并辅以过错推定原则，进而实现司法对于生成式人工智能提供者责任的动态调整。

参考文献

[1] 於兴中，郑戈，丁晓东.“生成性人工智能”与法律：以 Chat GPT 为例[J].中国法律评论，2023(2)：1－20.
[2] 朱荣荣.类 Chat GPT 生成式人工智能对个人信息保护的挑战及应对[J].重庆大学学报(社会科学版)，2023(9)：1－14.
[3] 陈兵.促进生成式人工智能规范发展的法治考量及实践架构：兼评《生成式人工智能服务管理暂行办法》相关条款[J].中国应用法学，2023(4)：108－125.
[4] 朱家珺.生成式人工智能虚假有害信息规制的挑战与应对：以 Chat GPT 的应用为引[J].比较法研究：2023(5)：34－54.
[5] 丛立先，李泳霖.生成式 AI 的作品认定与版权归属：以 Chat GPT 的作品应用场景为例[J].山东大学学报(哲学社会科学版)，2023(4)：171－181.
[6] 王迁.再论人工智能生成的内容在著作权法中的定性[J].政法论坛，2023(4)：16－33.
[7] 黄玉烨，刘云开.Chat GPT 热背后的冷思考[N].中国知识产权报，2023－03－31(11).
[8] 刘霜，张潇月.生成式人工智能数据风险的法律保护与规制研究：以 ChatGPT 潜在数据风险为例[J].贵州大学学报(社会科学版)，2023，41(5)：87－97.
[9] 王迁.知识产权法教程[M].第7版.北京：中国人民大学出版社，2021：82.
[10] 李艾真.美国人工智能生成物著作权保护的探索及启示[J].电子知识产权，2020(11)：81－92.
[11] 徐家力.人工智能生成物的著作权归属[J].暨南学报(哲学社会科学版)，2023(4)：37－49.
[12] 邓文.以 Chat GPT 为代表的生成式 AI 内容发可版权性研究[J].政治与法律，2023(9)：84－97.
[13] 韩大元.当代科技发展的宪法界限[J].法治现代化研究，2018，2(5)：1－12.
[14] 美国法律研究院.侵权法重述第三版：产品责任[M].肖永平，等，译.北京：法律出版社，2006.
[15] 周学峰.生成式人工智能侵权责任探析[J].比较法研究，2023(4)：117－131.

[16] 朱振.归责何以可能:人工智能时代的自由意志与法律责任[J].比较法研究,2022(1):39-54.
[17] ABBOTT R B. The Reasonable Computer: Disrupting the Paradigm of Tort Liability 86 GEO. WASH. L. REV. 1 (2018)[J]. SSRN Electronic Journal, 2016, 86(1):8.
[18] 邵红红.破解算法侵权责任界定的中立性难题:以"算法推荐第一案"为切入点[J].新闻界,2023(10):1-12.
[19] 何丽新,彭凯,刘静怡.搜索引擎"算法侵权"的归责路径探析[J].西北工业大学学报(社会科学版),2020(2):90-93.

Legal Risks and Infringement Liability Determination of Generative Artificial Intelligence

YAN Yaru, LUO Xiaochun
(School of Law and Politics, Nanjing Tech University, Jiangsu, Nanjing, 211816, China)

Abstract: The rapid development of generative artificial intelligence reflects the dialectical relationship between technological progress and the development of legal system. The current generative artificial intelligence technology brings numerous technological dividends, along with many legal risks, mainly in the fields of personal information protection, copyright issues of generated content, and the selection of infringement liability determining systems. Adhering to the principle of encouraging creativity and protecting the rights and interests of creators, we suggest that the content created by generative artificial intelligence should be recognized as copyrightable under certain circumstances, which provides a basis for the ownership of copyright and the determination of infringing parties. Taking into account the risk control capabilities and security obligation of each party, we suggest that we should adopt a general infringement liability determination system, supplemented by fault presumption rules, in consideration of the infringement liability of generative artificial intelligence. Based on the particularity of infringement of generative artificial intelligence, we should flexibly apply the "notice-and-takedown regime" on service providers.

Key words: generative artificial intelligence; copyrightability; products liability; fault liability principle; fault presumption

〔责任编辑:朱　根〕