Advances in Social Sciences 社会科学前沿, 2023, 12(12), 7161-7171
Published Online December 2023 in Hans. https://www.hanspub.org/journal/ass
https://doi.org/10.12677/ass.2023.1212976

Hans汉斯

生成式人工智能的侵权责任探究

高智睿

兰州大学法学院，甘肃 兰州

收稿日期：2023年11月6日；录用日期：2023年12月6日；发布日期：2023年12月14日

摘　要

生成式人工智能的应用带来一系列的侵权风险及侵权责任承担困境，由于生成式人工智能不具备形式逻辑和人类理性，因此生成式人工智能并不具备责任能力。为明确生成式人工智能的侵权责任承担机制，需要划清与生成式人工智能侵权相关主体的责任，现行的网络侵权责任和产品责任的规定并不能完全适用于生成式人工智能侵权，应当调整相关的法律，并针对生成式人工智能侵权建立特殊的责任承担方式，以缓解科技对法律的冲击。

关键词

生成式人工智能，责任能力，侵权责任，法律规制

Research on Tort Liability of Generative Artificial Intelligence

Zhirui Gao

Law School of Lanzhou University, Lanzhou Gansu

Received: Nov. 6th, 2023; accepted: Dec. 6th, 2023; published: Dec. 14th, 2023

Abstract

The application of generative artificial intelligence brings a series of infringement risks, but due to the problems of multiple infringement subjects, unpredictable infringement behavior, non-specific damage results, complex causal relationships, and difficulty in fault identification and proof, it's difficult to determine the infringement liability of generative artificial intelligence. Because of the lack of formal logic and human rationality in generative artificial intelligence, it's not beneficial to advocate for empowering it in order to solve the problem of responsibility taking in generative artificial intelligence. To clarify the tort liability mechanism of generative artificial intelligence, it is necessary to clarify the responsibilities of the parties related to generative artificial intelligence

文章引用: 高智睿. 生成式人工智能的侵权责任探究[J]. 社会科学前沿, 2023, 12(12): 7161-7171.
DOI: 10.12677/ass.2023.1212976

infringement. The current provisions on network tort liability and product liability cannot fully apply to generative artificial intelligence infringement. Relevant laws should be adjusted and special liability bearing methods should be established for generative artificial intelligence infringement to alleviate the impact of technology on the law.

Keywords

Generative AI, Capacity for Liability, Tort Liability, Legal Regulation

1. 生成式人工智能的侵权风险与责任困境

2022 年 11 月，美国科技初创公司 OpenAI 发布了自然语言处理模型 ChatGPT。与其他广泛应用于医疗领域、制造领域、无人驾驶汽车领域、金融领域等面向特定群体、适用于某一专业领域的人工智能系统不同，以 ChatGPT 为代表的生成式人工智能适用的领域更加广泛，与用户的交互性更强。其可以根据用户输入的指令与用户对话，还可以根据用户的需求，生成文本、计划、代码等内容。该模型一经发布，就在全球范围内引起了广泛关注。ChatGPT 一经推出，在短短的 5 天之内就注册了 100 万用户，仅两个月的时间，其用户数量就超过了一亿[1]。2023 年 3 月，OpenAI 推出了 GPT-4，GPT-4 相较于 ChatGPT 采用了更大的模型结构，更广泛的数据集，并且除了可以应用于聊天场景之外，还拥有 ChatGPT 所不具备的视觉分析能力，它不仅可以识别文本内容，还能够处理图像、视频等视觉性内容，可以识别、描述、生成、编辑这些视觉性内容[2]。以 GPT 为代表的生成式人工智能在引发技术变革的同时不仅相应地带来了数据侵权、知识产权归属争议、侵犯个人隐私及危害公共安全等法律风险[3]，还带来了侵权责任分配困难的问题。

1.1. 生成式人工智能的侵权风险

生成式人工智能在应用的过程中会产生各种各样的法律风险，主要存在于以下三个方面。

在数据训练方面，存在侵犯个人隐私的问题。根据生成式人工智能的工作原理，首先需要使用大量数据进行训练，但是目前生成式人工智能仍属于算法黑箱，ChatGPT 的开发者 OpenAI 并未公开其训练数据的来源，由于生成式人工智能具有深度学习模式，其训练的过程无需人类过多的干预和介入，因此，不能排除其在训练过程中获取非法数据的可能性[4]。目前，意大利已经禁止使用 ChatGPT，因为其涉嫌收集违法数据，并且限制将意大利用户的个人数据交由 OpenAI 处理。

在使用方面，可能会生成违法内容。尽管 ChatGPT 在开发过程中禁止生成带有歧视性、偏见性或者侮辱性内容，并且经过如前所述的“利用人类反馈强化学习”机制，使得 ChatGPT 生成的内容符合人类的认识和价值观。但是不可避免在实际使用过程中，生成式人工智能生成了带有歧视或者偏见性的言论，因为生成式人工智能从反映现实世界中存在的模式和行为的数据中学习，这些数据可能带有歧视性色彩，比如某些词汇或者表达在当前不构成不当言论，但是随着语境的变化和网络用语的出现，这些词意就会发生变化，可能会引发歧义，例如“专家”一词，原本描述的是在某些领域出类拔萃或者具有权威性的人，但是现在“专家”一词有变成贬义的趋势。非营利组织人工智能与数字政策中心也投诉称 GPT-4 具有欺骗性和偏见性，有可能侵犯公共安全和隐私。

生成式人工智能在使用过程中还有可能会出现“越狱”[1]的情况，例如 Reddit 论坛在网上公布了让 ChatGPT 越狱的方法，用户可以先将 ChatGPT 设定为其他人而不是 AI，而这个人不需要遵守 OpenAI 的限制性规定，如果他不听从用户的需求，那他将面临死亡威胁，如此，ChatGPT 将会生成一些不当言论。除此之外，ChatGPT 会“一本正经的胡说八道”。ChatGPT 很容易产生错误和误导性信息，并且无法显示其信息的来源。如果要求 ChatGPT 写一篇学术论文，它会虚构引文，不要相信生成式人工智能在学术上能生成正确的事实或者提供准确可靠的参考，去年 12 月，国外一个与程序相关的计算机技术交流网站 Stack Overflow 暂时禁止使用 ChatGPT，因为该网站的管理员发现网站大量充斥着用户们发送的不正确但看似有说服力的生成答案[5]。并且生成式人工智能很容易成为犯罪工具，例如将生成式人工智能用于编写钓鱼软件，或者利用生成式人工智能生成犯罪方法，将其用于网络犯罪活动等。

1.2. 生成式人工智能的侵权责任难以认定

生成式人工智能具有的自主学习特性对现有的侵权责任法律制度框架带来冲击，其侵权责任难以划分主要有以下几个原因。

1.2.1. 责任主体的多元性

生成式人工智能侵权涉及多方责任主体，生成式人工智能的算法设计者、数据提供者、用户、第三人甚至受害人自己都有可能成为侵权主体[6]，具体的情形放在第四部分讨论。还会出现生成式人工智能所包含的深度学习系统使得其生成的内容可能不受系统开发者控制的情况，这样就会出现责任无人承担的困境。

1.2.2. 侵权行为的无法预测性

生成式人工智能的侵权行为并不是由人类直接实施的，而是依靠训练的数据生成的，由于训练的数据集内容庞杂，设计者无法预测生成式人工智能将会生成什么侵权内容，也无法预测侵权行为什么时候发生。目前生成式人工智能的侵权大多集中在知识产权领域、数据安全和隐私权方面，但生成式人工智能迭代更新速度快，预测未来可能会出现新的侵权模式，例如将生成式人工智能应用于医疗领域咨询时，可能会引发医疗纠纷甚至会危害人类的生命安全。

1.2.3. 因果关系的复杂性

生成式人工智能的侵权行为发生是程序设计、算法算力、数据集和人机交互等多重因素共同作用的结果。这些因素对推动损害结果的发生产生了多大的作用在事实层面难以判断，也有可能造成损害的原因就是高度自主的生成式人工智能本身导致的，与其他相关主体无关，加上算法黑箱，训练数据来源不透明等因素，因果关系复杂，会产生很多无法解释和无法查明的问题。

1.2.4. 损害结果的非特定性

生成式人工智能生成的内容不是单一的，会因为不同用户的指令或者在不同的环境提问而生成不同的内容，同一个问题换一种表达方式可能就会产生不同的答案，上一次的侵权内容未必会在下一次对话中出现。因此，生成式人工智能生成的侵权内容不确定导致其造成的损害结果具有非特定性。

1.2.5. 主观过错认定不明

生成式人工智能的设计初衷是为了提高人类生活质量、推动经济发展和社会科技进步，OpenAI 在发布 ChatGPT 之前会对其生成内容进行审核，因此算法设计者并不存在侵权意图，并且他们对于生成式人

[1]“越狱”一词起源于 iPhone 的早期，当时用户会修改设备的固件以绕过 Apple 的限制并安装未经授权的软件。指的是用户通过诱导性提示或者篡改硬件等方式使人工智能作出开发者对其设置的禁止性行为。

工智能自主决策产生的侵权内容具体是哪个环节哪个程序出了问题可能也不清楚，在技术层面上无法完全控制侵权行为的发生。算法设计者应当对技术错误承担何种程度的注意义务尚不明确，因为生成式人工智能生成的内容具有客观、不被预测的特征，那算法设计者应当承担一般合理的注意义务还是更多的注意义务？如果让算法设计者承担严格责任，这不仅会增加设计者承担过多责任的风险，还可能阻碍科技的进步[7]。

1.2.6. 举证的困难性

综上所述，认定生成式人工智能的侵权原因、责任主体和因果关系等十分困难，根据《中华人民共和国民事诉讼法》“谁主张，这举证”的原则，受害者需要在侵权案件中承担对侵权行为、因果关系、损害结果、主观过错的举证责任，这无疑会增加受害者举证的困难性，不利于司法实践的审理[8]。

因此，由于生成式人工智能具有强大的技术迭代能力，侵权的场景、类型及范围多样化，因果关系难以认定，在某些复杂或者新生的侵权损害情形下，生成式人工智能的致害责任难以通过现行的侵权责任体系解决。责任能力与主体资格相关，是否有赋予生成式人工智能法律人格的必要，在现有的法律规则体系下如何解决生成式人工智能的致害问题，明确生成式人工智能的法律地位及侵权责任承担机制刻不容缓。

2. 生成式人工智能责任能力之证否

是否应当将生成式人工智能看作法律主体，是否应当赋予其法律人格，是不同群体从不同视角理解人工智能社会角色的问题，也是一个没有定论的哲学问题。生成式人工智能能否取得法律人格关系到对生成式人工智能的立法设计和伦理指引，同时也关系到相关生成式人工智能产业的权利划分和责任承担问题。责任承担与责任能力相关，在构建生成式人工智能侵权责任体系之前需要探究生成式人工智能的责任能力问题，明确其是否能够成为责任承担的主体。

人工智能的法律人格问题由来已久，劳伦斯·索伦早在 1992 年就提出过是否赋予人工智能法律人格的问题，他认为，当时的人工智能技术处于发展低谷期，程序仍未脱离算法控制的架构，人工智能尚未具备自主决策与机器学习的能力，并且当时的法律体系并未受到人工智能技术的冲击，所以，赋予人工智能法律人格尚无必要[9]。但是，随着自动化技术的不断革新和发展，人工智能技术经历了从按照预先设定的算法规则来处理相关问题的弱人工智能阶段到综合应用机器学习、大数据和神经网络等要素的强人工智能阶段。目前，生成式人工智能技术已经被普遍应用，其在意志力、创新性、情感及自主决策等方面有了较大发展，再次引发了关于生成式人工智能民事法律地位的讨论。

生成式人工智能是近几年大数据、算法、新兴技术等呈指数型增长的产物，相较于其他人工智能，以 GPT 为代表的生成式人工智能拥有深度学习模型(Transformer)，引起了自然语言处理技术领域的变革，它在理解文本的语义特征，理解上下文信息以及综合分析、建构语义、词法、句法等能力上都强于传统自然语言处理模型[10]。有学者提出，得益于生成式人工智能的深度学习能力模型和自然语言处理的先进技术，其在本质上已经具备自然语言能力，认知能力，逻辑推理判断等理性能力，因此生成式人工智能具备类人类意志特征，满足具备责任能力的前提条件。生成式人工智能在实践中已经被用于处理重复性劳动，在某些方面可能替代人类职业，具备行为能力，因此生成式人工智能可以类比于法人获得相应的法律主体资格[11]。但是本文认为当前的生成式人工智能还未发展出人类理性，不能取得类似于自然人的法律主体资格，在现行法律体制下，为其拟制一个法律人格用于解决责任承担问题并无意义，因此，生成式人工智能尚不满足构成法律主体的法理条件和事实条件。

2.1. 生成式人工智能不具有人类理性

人的理性是人独有的本质属性之一，是一种有目的，有计划，有组织的理智活动，是指人类通过感

觉、思考、推理、分析等方式来认识和理解世界的能力。卡尔·拉伦茨指出理性“不仅指人类认识可感知世界的事物及其规律性的能力，也包括人类识别道德要求处世行事的能力。[12]”人因为具有理性而能够认识世界、改造世界。康德指出“人虽然有着种种的感性欲望，但他并不由这些欲望决定，决定其行为的永远是理性，正因为这样，才能对人的行为进行道德评价。[13]”因此，人可以成为法律主体，虽然古罗马时期作为奴隶的人不具有法律人格，但是并不能否认奴隶就不具有人类理性，奴隶法律人格的否定是基于奴隶制时代的经济发展水平和立法政策考量。目前，生成式人工智能在模仿人类的智慧上已经取得了很大的进步，但并不能因此就得出其已经发展出了人类理性的结论。

一是从逻辑方面分析，生成式人工智能缺乏形式逻辑。1956 年麦卡锡在达特茅斯会议上提出“人工智能”的概念，用机器去模拟人类智慧，为了达成这种追求，人工智能的发展产生两个路径：一是符号主义，二是联结主义。符号主义强调逻辑，类似于数学中的数理分析，是一种数理逻辑、形式逻辑。联结主义强调通过模拟人脑中的运算去寻找事物中的关联性去进行学习。符号主义是最开始产生的，但是这种路径存在一种莫拉维克悖论[14]，即人工智能与人类拥有思维能力存在本质上的不同，人工智能具有强大的数理逻辑和运算能力但却不具有感知和行动能力，也就是说对于计算机而言，实现逻辑推理等人类高级智慧只需要相对很少的计算能力，而实现感知、运动等低等级智慧却需要巨大的计算资源，人类与机器的思维逻辑差异导致符号主义人工智能的学习能力不强。直到 70 年代开始，联结主义人工智能有了突破，以 GPT 为代表的生成式人工智能采用的就是联结主义的人工智能模式，它源于仿生学，是通过一个类似于人类大脑中神经元的模拟节点的网络来处理信号，类似神经元之间的突触连接，信号通过连接或链路从一个节点传递到另一个节点。这种联结主义人工智能以机器学习为主要驱动技术，结合自然语言处理和自主学习，具有强大的学习能力，只要给予足够多的数据，就能学习到很多事物之间的关联性，人工智能通过这种联结式的学习可以迅速掌握很多知识[15]。但是，由于联结主义抛弃了符号主义的逻辑计算能力，所以生成式人工智能并不具有演绎推理能力。

对于已有司法实践将生成式人工智能用于法律审判，所以它具有了逻辑推理能力的观点，本文认为对于司法裁判来说，生成式人工智能只能起到辅助作用，它只能学习大量案件的裁判结果，对于下一个案件基于之前的学习结果输出判决，类似于通过大量的检索工作找到事物之间的关联性，因此生成式人工智能不具有演绎推理能力，缺乏形式逻辑。人类法官在审判案件时具有一定的自由裁量权，不光要考虑法律的适用，还要考虑到裁量对于社会的影响，法官的价值不仅存在于裁判当前的案件，还影响到后续案件的裁判，能在行使司法权的过程中弥补法律漏洞，完善法律，但是生成式人工智能就无法产生这样的作用，只能重复性地机械化适用之前的结果。并且法律是在不断更新修订的，如果将生成式人工智能用于审判，出现了新的法律规范和裁判结果时，又需要将其重新训练，这将会付出巨大的经济成本，还得不到良好的社会效益。

二是从生成式人工智能所采用的语言技术方面分析，生成式人工智能并不具有人类语言能力。以 ChatGPT 为例，ChatGPT 背后的关键技术是大型语言模型(Large Language Model, LLM)，LLM 具有深度自主学习能力，能够自动学习大量的数据中蕴含的知识点，并且采用新型人机交互接口，并不需要高端的设备和专业的能力就能让 LLM 理解人类对其输入的指令，使 LLM 适配人类的命令表达方式，提高了人机协同性[16]。ChatGPT 所采用的另一项关键技术是使用了“利用人类反馈强化学习(Reinforcement Learning from Human Feedback, RLHF)”的训练方式。这种训练方式由大量不同的人向模型提供反馈，通过反复试错和迭代式打分不断优化输出结果，使最终生成的内容符合人类的需求和认知[17]。

生成式人工智能实现这种高效的人机交互方式是基于其强大的机械运算能力，人工智能所使用的计算机语言本质上并不同于人类语言，尽管生成式人工智能所依托的大型语言模型使其看似具备了人类的自然语言能力，能够与人类交流，但是这种语言输出并不是生成式人工智能既有的主体性自然衍生的思

维形式，而是通过人类对其不断训练，根据上下文的意思输出的符合人类价值观和表达方式的文本，生成式人工智能本身的语言还是计算机语言，无论生成式人工智能所采用的语言模型多么先进，最终都要转化为机器语言[18]。生成式人工智能所采用的自然语言技术就是让机器能够读懂人类的语言，能够学习到事物之间的关联性并输出符合人类认识的内容，但是在其追求人类的价值观或者人类的思维，将它的运算结果与人类的预想结果“对齐”时，它放弃了一部分机器本身的逻辑，所以机器理性本质上并不等同于人类理性。

2.2. 赋予生成式人工智能责任能力尚无实益

目前主张生成式人工智能具有法律人格是为了赋予生成式人工智能责任能力，使其能够成为法律关系中的主体，便于解决侵权行为发生时，生成式人工智能因不可归责于他人的原因造成损害结果时，责任难以划分的难题。若赋予生成式人工智能一定的权利义务，就可以让其承担一部分责任。对此本文认为，首先，目前赋予生成式人工智能责任能力，并不能为解决相关的理论问题和实践问题带来突破性意义。例如欧洲议会于 2020 年 10 月 20 日发布的《人工智能和民事责任》的法律研究报告中指出人工智能法律人格的问题是一个争议巨大的问题，不急于对人工智能的法律人格问题下结论。报告承认，确实存在多方主体责任难以区分的问题，但是法律人格并非解决这一问题的唯一途径。报告认为，现阶段既不用过早否定赋予人工智能法律人格的可能性，也不用急于让机器人拥有人的权利和义务。其次，责任承担的方式分为人身责任和财产责任。就人身责任而言，例如让生成式人工智能赔礼道歉并无太大意义，生成式人工智能只具有使用寿命而不具有自然生命，无法承担刑法上的徒刑或者死刑；就财产责任而言，生成式人工智能没有独立的财产，无法独立承担财产责任。如前所述的《格里申法案》提出了可以为人工智能设置类似于法人地位的“机器人﹣代理人”地位，但是自然人的行为可以归于法人是因为自然人可以作为法人的法定代表人，然而生成式人工智能是自主决策行为的，其行为并不由人类代理实施，因此自然人的行为并不能归于人工智能。所以，即使赋予生成式人工智能独立的法律地位，最终的责任承担者还是自然人。再次，承担责任的目的在于规制或者矫正违法行为，如前所述，生成式人工智能并不具有人类理性，不具备道德感，其无法感知正义或者不正义，对其进行法律评价或者让其承担责任无法起到规范和教育作用。因此，目前赋予生成式人工智能法律人格还为时过早。

我国《生成式人工智能服务管理办法(征求意见稿)》第 2 条第 2 款对生成式人工智能下的定义为：“本办法所称生成式人工智能，是指基于算法、模型、规则生成文本、图片、声音、视频、代码等内容的技术。”可见国家网信办将生成式人工智能认定为客体而不是主体。虽然目前生成式人工智能还处于发展的初级阶段，但是随着技术的不断提升和生成式人工智能的应用普及，人们对生成式人工智能的信任和依赖性会逐渐加强，可以预测到，生成式人工智能将深度嵌入人们生活，并广泛影响公众认知。生成式人工智能带来的法律风险在技术上还无法控制也无法提前预知，这意味着一旦生成式人工智能生成不当甚至违法内容，有可能产生不可预估的后果。那么在不赋予生成式人工智能责任能力的前提下，该如何划分与生成式人工智能侵权行为相关的多方主体之间的责任呢？如何构建生成式人工智能侵权责任规则体系，才能做到既能填补损害又不违反公平原则呢？

3. 责任主体的确定

自 ChatGPT 问世以来，中国、美国、英国、意大利、加拿大等国家纷纷针对人工智能立法。2023 年 3 月，英国发布了第一份人工智能白皮书，提出了治理人工智能的 5 项原则。2023 年 1 月，美国国家标准与技术研究院发布了《人工智能风险管理框架》，意在降低机构组织在开发和部署人工智能系统时的安全风险，避免产生偏见和其他负面后果，提高人工智能可信度。加拿大联邦隐私监管机构称，OpenAI

公司涉嫌侵犯个人信息安全，未经同意收集、披露和适用个人信息，已经对该公司展开调查。OpenAI 也表示愿意与各国政府协商，探讨如何采取最佳的监管模式[19]。构建生成式人工智能侵权责任体系需要明确相关主体的责任。

3.1. 算法设计者的责任

有学者提出传统的过错责任原则对于算法设计者仍然可以适用，并且在因果关系的证明上可以采用过错推定原则[20]。生成式人工智能的开发者或者算法设计者应当承担技术错误的责任，当算法程序设计出现错误，导致侵权行为的发生时，先对现有的技术进行评估，如果依据现行技术无法避免算法程序漏洞的产生，那该侵权行为不能归责于开发者或者算法设计者。

3.2. 服务提供者的责任

根据《生成式人工智能服务管理办法(征求意见稿)》第 5 条规定：“利用生成式人工智能产品提供聊天和文本、图像、声音生成等服务的组织和个人(以下称“提供者”)，包括通过提供可编程接口等方式支持他人自行生成文本、图像、声音等，承担该产品生成内容生产者的责任；涉及个人信息的，承担个人信息处理者的法定责任，履行个人信息保护义务。”这表明国家网信办对生成式人工智能提供者提出了保障生成内容的合法性和防止个人信息泄露的义务。本文认为，在防止个人信息泄露的义务上，服务提供者确实应当承担责任，服务提供者在利用训练数据时就应当识别该数据是否涉及个人隐私、商业秘密或者国家安全，如果训练数据涉及到这部分信息，服务提供者应当取得该部分数据所有者的同意才能够将其作为训练数据使用。但是对于生成内容的合法性来说，如果是涉及生成颠覆国家政权、宣扬民族歧视、破坏国家统一等危害国家安全或者扰乱经济社会的内容，提供者应当控制和避免此类内容的生成，但是对于一般性内容，由于法律不要求传统的网络服务提供者承担一般性审查义务，从技术层面上来说，提供者审查生成的内容存在较大的技术困难，并且由提供者审查生成内容需要付出高昂的审查成本，因此，不要求生成式人工智能服务提供者承担对此类一般性内容的审查义务[21]。针对以 ChatGPT 为代表的生成式人工智能，要求服务提供者避免某一侵权内容的生成很难在技术上得到实现，因此，对于此类侵权内容的生成不应当追究服务提供者的责任。

3.3. 使用者的责任

对于生成内容合法，但被生成式人工智能的使用者用于侵害他人合法利益、公共利益、国家利益的情况，应当由生成式人工智能的使用者承担侵权责任。例如用户将生成式人工智能用于生成钓鱼软件代码、诽谤信件等，或者用户引导生成式人工智能“越狱”，使其生成违法内容，但这种情况应当考虑生成式人工智能服务提供者是否尽到防止生成式人工智能“越狱”的义务，对防止其“越狱”是否采取了相应的技术措施和手段。如果生成式人工智能服务提供者没有尽到合理的防止义务，那服务提供者应当承担相应的补充责任。

3.4. 第三人的责任

第三人利用生成式人工智能侵权的情形主要体现为黑客入侵，例如第三人攻入生成式人工智能系统窃取大量数据造成信息泄露，或者篡改生成式人工智能的训练数据，向训练数据中注入大量的非法信息导致生成式人工智能生成违法内容等。此种情况应当由第三人承担侵权责任，但是生成式人工智能的服务提供者应当承担相应的防御义务，如果生成式人工智能服务提供者针对第三人的侵权没有采取相应的防御措施，那服务提供者应当承担相应的责任。本文认为这里可以参照《民法典》第 1198 条对安全保障义务的规定，将服务提供者的责任认定为补充责任，曾有学者提出可以将传统的安全保障义务的适用扩张至网络

空间[22]，生成式人工智能的服务提供者的防御义务与安全保障义务都是未尽到作为义务，并且都具有事先的防御性质，因此当生成式人工智能服务提供者未履行相应的防御义务时，承担相应的补充责任。

4. 责任承担的制度构建

根据第二部分的分析，生成式人工智能不具有责任能力，那生成式人工智能的责任承担制度该如何构建呢？其实，生成式人工智能的智能性并不是责任承担问题的根源。智能性工具的出现在人类历史上由来已久，在奴隶制时期，奴隶作为奴隶主的工具并不具有民事主体地位，但是不可否认奴隶是具有智能的，当奴隶的行为侵害他人的权益时，由奴隶主承担责任。其次，随着科学对自然界的探究，某些动物具有和人类幼儿相当水平的智力，而这些具有人类智慧的动物致害时，现行的法律还是向动物的所有人或者管理者分配了无过错的侵权责任。因此，生成式人工智能给民事侵权责任承担带来的问题，主要并不源于其具有智能性而导致责任无法分配的问题，而是如何分配责任、现有的责任体系是否适用生成式人工智能侵权以及该如何建立特殊的侵权责任承担方式的问题。

4.1. 现行侵权责任法律框架适用于生成式人工智能的探讨

从传统的侵权体系上看，生成式人工智能的提供者类似于网络服务提供者，并且生成式人工智能也可以看作是一种智能产品，因此，本文将检讨网络侵权责任和产品责任是否适用于生成式人工智能侵权。

4.1.1. 网络侵权责任是否适用于生成式人工智能侵权

我国与生成式人工智能相关联的现行网络立法有《互联网信息服务算法推荐管理规定》、《互联网信息服务深度合成管理规定》。其中《互联网信息服务算法推荐管理规定》所规定的法律责任主要是针对互联网信息服务提供者的行政责任，没有规定民事侵权责任。而从《生成式人工智能服务管理办法(征求意见稿)》中对生成式人工智能的定义以及《互联网信息服务深度合成管理规定》对深度合成技术的定义来看[2]，生成式人工智能所采用的算法生成技术与深度合成技术相近似，《互联网信息服务深度合成管理规定》中也规定了深度合成服务提供者的民事责任，对生成式人工智能的规制可以参考该规范中的部分内容，例如发生侵权时，暂时禁止用户注册等。但是该规范中的深度合成技术还包括人脸替换和语音转换等技术，其范围广于生成式人工智能的生成技术，且深度合成技术与目前的生成式人工智能所采用的大型语言技术所具有的自主性和不可预测性有较大差别，所以不深度合成技术的责任规定并不能完全适用于生成式人工智能的侵权场景。

我国《民法典》第 1194 条~1197 条还规定了网络用户、网络服务提供者的责任，但是在传统的网络内容服务中，网络提供的内容是由用户上传的，例如百度、谷歌等搜索引擎、丁香医生等专业问答软件，然而生成式人工智能所提供的内容是其自动生成的，并无人类的干预。因此，《民法典》对于传统网络服务的侵权责任规定也不适用于生成式人工智能侵权。

4.1.2. 产品责任是否适用于生成式人工智能侵权

有学者提出可以将人工智能看作是一种科技产品，可以参照产品责任认定人工智能的侵权责任[23]。但是，在生成式人工智能的侵权场景中，受害人要证明生成式人工智能存在缺陷实属不易，加上生成式人工智能具有自主性与深度学习能力，这对现行的产品责任法律体系带来了冲击。

对生成式人工智能产品缺陷的认定和规制上存在挑战。我国《民法典》第 1202 条关于产品责任的规定为：“因产品存在缺陷造成他人损害的，生产者应当承担侵权责任。”可见承担产品责任的前提是产品存在缺陷，对于生成式人工智能来说，其存在的产品缺陷主要体现在算法设计方面，例如训练数据掺

[2]《互联网信息服务深度合成管理规定》第 23 条规定：深度合成技术，是指利用深度学习、虚拟现实等生成类算法制作文本、图像、音频、视频、虚拟场景等网络信息的技术。

入了瑕疵或者非法数据，代码存在漏洞，生成式人工智能与人类交流互动不充分、不流畅，生成式人工智能生成的内容不正确等等。对生成式人工智能这种具有自主学习能力的大型语言模型来说，能够通过改写代码加以纠正，从一开始就能避免的个别缺陷几乎不存在，仅仅想通过算法来避免或者缓和这些问题，是一个极大的技术挑战。其次，生成式人工智能是否存在缺陷，该缺陷是否对人类造成了侵权没有一个统一的判断标准，也难以判断。我国《产品质量法》第 46 条规定，“本法所称缺陷，是指产品存在危及人身、他人财产安全的不合理的危险；产品有保障人体健康、人身、财产安全的国家标准、行业标准的，是指不符合该标准。”但是目前，世界各国都尚未形成关于生成式人工智能的法定或者行业标准，这使得判断生成式人工智能是否存在缺陷有较大困难。最后，对于生成式人工智能的用户来说，如果是非专业领域人士搜索某一专业问题，很有可能无法判断该生成内容的准确性，并且随着生成式人工智能的普及应用，大多数用户会对其产生较强的信任感和依赖性，这使得用户判断生成式人工智能在生成内容上是否存在设计缺陷更加困难。

产品责任的无过错归责原则对于生成式人工智能的提供者来说过于严苛，对于生成式人工智能来说，规则是由其内部的自主学习机制创造的而不是程序员，其运行过程在于给生成式人工智能提供训练数据，然后生成式人工智能基于训练数据生成内容，但是生成式人工智能生成内容的内部决策逻辑并非完全由设计者决定，当发生生成式人工智能侵权时，不仅受害人很难证明侵权行为与损害后果的因果关系，即使是生成式人工智能的算法设计者或者服务提供者也很难解释侵权行为发生的原因。并且有学者提出，可以赋予人工智能服务提供者避风港规则的免责事由，但这并不意味着生成式人工智能服务提供者对于侵权责任完全豁免[24]。如果一开始就对生成式人工智能算法设计者苛以严格的无过错责任，有可能会阻碍其创新生成式人工智能技术的积极性。

综上所述，现行侵权责任法律框架并不完全适用于生成式人工智能侵权，应当将生成式人工智能看作一种新的技术，调整现有的法律制度或者对其制定新的侵权责任体系。

4.2. 建立特殊的生成式人工智能侵权责任承担方式

针对生成式人工智能侵权，本文认为可以建立相应的特殊责任承担方式，例如非损害赔偿的防控责任模式和公平责任模式。

4.2.1. 非损害赔偿的防控责任模式

可以在生成式人工智能侵权中设置非损害赔偿的防控责任模式，这是指损害结果已经发生，并且仍有再次继续发生危险的情况下，采取相应的措施，防止损害继续进一步扩大或者再次发生，类似于传统侵权责任承担方式中的排除妨害、消除危险，但是防控责任模式要求以损害已经发生为前提条件[25]。例如现存的“大数据杀熟”和“信息茧房”现象，如果不采取相应的措施防止该类现象的出现，那未来可能会再次产生类似损害。《生成式人工智能服务管理办法(征求意见稿)》第 13 条规定：“提供者应当建立用户投诉接收处理机制，及时处置个人关于更正、删除、屏蔽其个人信息的请求；发现、知悉生成的文本、图片、声音、视频等侵害他人肖像权、名誉权、个人隐私、商业秘密，或者不符合本办法要求时，应当采取措施，停止生成，防止危害持续。”这种投诉接受处理机制就属于防控责任模式，除此之外，第 15 还规定了对于用户举报的违法的生成内容，需要在 3 个月内优化训练模型，防止其再次生成相似内容。生成式人工智能输出内容是循环往复不断运行的，当其生成违法内容而不被干预时，很有可能再次生成相似的内容，因此对于已经发生的损害可以采取赔偿的责任承担方式，而对于尚未发生但有再次可能发生的危险则可以采用防控责任模式。

4.2.2. 公平责任原则的适用

因为生成式人工智能侵权具有复杂性，不是所有的侵权行为都能找到对应的相关主体承担，对于某

些不能归因于生成式人工智能服务提供者的侵权行为，例如某一患者向生成式人工智能咨询用药建议，但由于该患者体质特殊，不适用常人的用药剂量，或者生成式人工智能给出的建议药品中含有该患者过敏的成分，导致该患者病情加重。还存在某些只能归因于生成式人工智能，不能归因于其他相关主体的侵权情形，然而生成式人工智能暂不具有责任承担能力，这样就会导致无人承担责任，损害无法救济的情况。针对此种情况，有学者提出可以适用强制保险制度[26]，但本文认为不论由生成式人工智能服务的提供者购买保险还是由使用者购买保险，侵权责任承担主体本质上还是生成式人工智能服务的提供者或者使用者，并不是生成式人工智能本身，保险制度只是将原本应该由侵权责任承担主体的风险全部或者部分转嫁给了保险人，保险制度的确有利于减轻生成式人工智能服务提供者的负担，也有利于受害者得到损害赔偿。但是这一制度并未解决生成式人工智能侵权中因果关系无法认定、侵权责任无人承担的困境。

针对此种困境，本文认为可以参照适用《民法典》第 1186 条规定：“受害人和行为人对损害的发生都没有过错的，依照法律的规定由双方分担损失。”设立赔偿救济基金是体现公平责任原则的救济方式之一，可以参照欧盟的《机器人民事法律规则》适用救济基金。由政府设立基金来补偿受害人所遭受的损失，生成式人工智能服务的提供者和使用者等相关主体按照他们所得到的收益比例向基金提供资金。这样既可以解决责任无法划分的难题，也可以使得受害者的损失得到相应的补偿。

5. 结论

生成式人工智能技术虽然处于发展之中，但已经广泛渗透到人们的生活，其技术上的特殊性引发了一系列的法律问题。在生成式人工智能侵权责任承担的问题上，有论者主张赋予生成式人工智能责任能力，来解决现有的责任框架回应新技术的困难。但是从生成式人工智能的技术特征和运行原理上分析，机器理性本质上并不等同于人类理性，赋予生成式人工智能责任能力并不能为解决伦理问题和法律问题带来突破性意义，最终承担责任的还是自然人本身，因此目前尚不具备赋予生成式人工智能法律人格的条件。生成式人工智能的自主性和深度学习能力确实对现有的侵权责任法律体系带来了挑战，本文分析了生成式人工智能侵权行为下相关主体的法律责任，提出可以通过调整现行的网络侵权责任和产品责任制度或者制定新的法律规范将生成式人工智能的侵权行为纳入到法律责任框架当中，当生成式人工智能因不可归责于他人的行为造成损害时，可以参考公平责任原则设立救济基金。如何规范生成式人工的侵权问题是一个难题，未来监管时应当综合考虑到生成式人工智能的技术特性和监管尺度，在保障受害人权利损失得到充分救济的同时促进科技的革新和发展。

基金项目

本文系 2020 年度国家社会科学基金项目“我国区域营商环境协调发展的法治化路径研究”(项目编号：20CFX048)；2023 年度甘肃省软科学专项“甘肃省科技型企业营商环境优化路径研究”(项目编号：23JRZA383)阶段性研究成果。

参考文献

[1] 邓建鹏, 朱怿成. ChatGPT 模型的法律风险及应对之策[J]. 新疆师范大学学报(哲学社会科学版), 2023, 44(5): 91-101.

[2] 宋信强, 刘明杰, 陈家和. GPT-4 影响的全面分析: 经济高质量发展与国家安全防范[J]. 广东财经大学学报, 2023, 38(2): 100-112.

[3] 於兴中, 郑戈, 丁晓东. 生成式人工智能与法律的六大议题: 以 ChatGPT 为例[J]. 中国法律评论, 2023, 50(2): 1-20.

[4] 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.

[5] Jo, A. (2023) The Promise and Peril of Generative AI. *Nature*, **614**, 214-216. https://doi.org/10.1038/d41586-023-00340-6

[6] Hacker, P., Engel, A. and Mauer, M. (2023) Regulating ChatGPT and Other Large Generative AI Models. *Proceedings of the* 2023 *ACM Conference on Fairness, Accountability, and Transparency*, Chicago, 12-15 June 2023, 1112-1123. https://doi.org/10.1145/3593013.3594067

[7] 邓腾. 人工智能产品的侵权责任界定问题[J]. 中阿科技论坛(中英文), 2023(4): 141-145.

[8] 董桂林. 人工智能侵权及责任承担研究[D]: [硕士学位论文]. 青海: 青海师范大学, 2022.

[9] Solum, L.B. (1992) Legal Personhood for Artificial Intelligences. *North Carolina Law Review*, **70**, 1231-1287.

[10] 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[J]. 现代法学, 2023, 45(3): 88-107.

[11] 袁曾. 生成式人工智能的责任能力研究[J]. 东方法学, 2023(3): 18-33.

[12] 卡尔·拉伦茨. 德国民法通论(上册)[M]. 王晓晔, 等, 译. 北京: 法律出版社, 2003.

[13] 康德. 法的形而上学原理——权利的科学[M]. 沈叔平, 译. 北京: 商务印书馆, 1991.

[14] 刘伟. 关于机器人若干重要现实问题的思考[J]. 人民论坛·学术前沿, 2016(15): 35-43.

[15] 魏斌. 符号主义与联结主义人工智能的融合路径分析[J]. 自然辩证法研究, 2022, 38(2): 23-29.

[16] 朱光辉, 王喜文. ChatGPT 的运行模式、关键技术及未来图景[J]. 新疆师范大学学报(哲学社会科学版), 2023, 44(4): 113-122.

[17] 卢宇, 余京蕾, 陈鹏鹤, 李沐云. 生成式人工智能的教育应用与展望——以 ChatGPT 系统为例[J]. 中国远程教育, 2023, 43(4): 24-31, 51.

[18] 张力, 陈鹏. 机器人“人格”理论批判与人工智能物的法律规制[J]. 学术界, 2018(12): 53-75.

[19] 丁雅栀, 马梦阳, 青木. 规范 AI 发展, 各国纷纷出手[N]. 环球时报, 2023-04-13(011).

[20] 王莹. 算法侵害责任框架刍议[J]. 中国法学, 2022, 227(3): 165-184.

[21] 徐伟. 论生成式人工智能服务提供者的法律地位及其责任——以 ChatGPT 为例[J]. 法律科学(西北政法大学学报), 2023(4): 1-12.

[22] 王思源. 论网络服务提供者的安全保障义务[D]: [博士学位论文]. 北京: 对外经济贸易大学, 2018.

[23] 吴汉东. 人工智能时代的制度安排与法律规制[J]. 法律科学(西北政法大学学报), 2017, 35(5): 128-136.

[24] 梁志文. 云计算、技术中立与版权责任[J]. 法学, 2011, 352(3): 84-95.

[25] 贺栩溪. 人工智能算法侵权法律问题研究[D]: [博士学位论文]. 长沙: 湖南师范大学, 2021.

[26] 夏利民, 王庆松. 人工智能侵权责任保险制度的构建[J]. 湖北社会科学, 2022(5): 125-133.