Open Journal of Legal Science 法学, 2024, 12(7), 4245-4251
Published Online July 2024 in Hans. https://www.hanspub.org/journal/ojls
https://doi.org/10.12677/ojls.2024.127602

Hans汉斯

生成式人工智能的法律风险与规制路径

盛梦雅

南京林业大学马克思主义学院，江苏 南京

收稿日期：2024年5月17日；录用日期：2024年6月5日；发布日期：2024年7月4日

摘　要

生成式人工智能的快速发展不仅对社会生活产生了多方面影响，也引发了一系列法律风险。基于生成式人工智能技术的起源和展望，重点介绍了所引发的数据安全问题、知识产权问题、虚假信息问题等法律风险，针对性地提出了面对风险的规避路径。具体包括加强数据监管，赋予数据主体必要的数据权利；完善法律法规，健全知识产权纠纷机制；优化内容生成技术，明确责任主体义务。

关键词

生成式人工智能，法律风险，法律规制

Legal Risks and Avoidance Paths of Generative Artificial Intelligence

Mengya Sheng

School of Marxism, Nanjing Forestry University, Nanjing Jiangsu

Received: May 17th, 2024; accepted: Jun. 5th, 2024; published: Jul. 4th, 2024

Abstract

The rapid development of generative artificial intelligence has not only affected social life in many ways, but also triggered a series of legal risks. Based on the origin and outlook of generative AI technology, it focuses on the legal risks arising from data security issues, intellectual property rights issues, false information issues, etc., and puts forward targeted avoidance paths to face the risks. Specifically, it includes strengthening data regulation and giving data subjects the necessary data rights; improving laws and regulations, and sounding the mechanism of intellectual property disputes; and optimizing content generation technology and clarifying the obligations of responsible subjects.

文章引用: 盛梦雅. 生成式人工智能的法律风险与规制路径[J]. 法学, 2024, 12(7): 4245-4251.
DOI: 10.12677/ojls.2024.127602

Keywords

Generative Artificial Intelligence (AI), Legal Risk, Legal Regulation

1. 引言

在 2022 年 11 月，美国 OpenAI 公司正式发布了名为“ChatGPT”的聊天机器人程序。与以前的人工智能相比，ChatGPT 能够理解并学习人类的语言进行即时地对话，并根据对话情境进行深度互动的特性，可以真正地与人类进行沟通交流，甚至可以完成撰写邮件、编写视频脚本、创作文案、翻译以及编程等工作。以 ChatGPT 为代表的新一代人工智能，即生成式人工智能(Generative AI)，与以往人工智能的主要目的是分类与回归不同，是一种可以创建各种数据，例如，图片、视频、音频、文本和 3D 模型的人工智能系统。

ChatGPT 上线仅 2 个月，已拥有超过 1 亿的用户，是有史以来用户量增长速度最快的应用程序。然而，伴随着 ChatGPT 的迅速普及，其引发的各类法律风险也逐渐显露出来，如数据侵权、虚假信息、算法歧视、侵犯隐私等问题。例如，ChatGPT 既能模仿真实数据并生成新的内容，又可能会被不法分子利用，制造虚假信息、误导公众或散播不实言论。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》)。办法作为我国出台的首部对生成式人工智能进行治理的行业法规，其目的在于对生成式人工智能进行全方位的治理和监管。此外，国外也在积极关注生成式人工智能的安全风险问题，2023 年 6 月，欧洲议会通过了 2021 年 4 月由欧盟委员会提出的《人工智能法案》，该法案将人工智能的风险划分为不同等级并通过区分策略进行普适性监管治理[1]。在此背景下，本文以生成式人工智能的应用为切入点，对其产生的一些常见的法律应用风险进行分析，并结合该风险构建生成式人工智能应用的法律规制路径。

2. 生成式人工智能的技术发展

伴随着人工智能(AI)技术的迅猛发展，生成式人工智能(GenAI)正在逐步崛起，成为推动新一轮科技革命和产业变革的重要驱动力。生成式人工智能作为颠覆性技术集成式创新的产物，它不仅塑造了人类社会生产和生活形态，还将人类社会推动迈向数字文明的新形态阶段[2]。只有充分理解技术发展的起源和关键发展阶段，展望未来发展趋势，才能提出有针对性的治理路径。

(一) 生成式人工智能的起源

生成式人工智能的起源最早可以追溯到上世纪 50 年代，当时人工智能领域的专家们就开始研究如何利用计算机生成新的、有用的信息，即提升算法模型智能化理解通用语言方面能力的尝试。传统的语言模型学习依赖于结构化数据对其进行建模和处理，这制约了其在无标注数据上的应用。另外，基于长短记忆模型构建的模型不能充分利用外部词汇，也不能将整个语料作为一个整体来处理，导致整个语言模型的训练速度受到限制。从本质上看，语言模型的文本处理能力通常由三个因素决定：计算量、模型参数的数量和训练数据集的大小[3]。伴随着深度学习和大数据等技术的兴起，生成式人工智能得到了迅速的发展，它以复杂的算法、模型和规则为基础，从大规模数据集中学习出具有原创性的新内容。生成式

人工智能的迅速发展得益于三个关键领域的人工智能技术，即生成算法、预训练模型和多模态技术。ChatGPT 基于 OpenAI 的 GPT3.5 自回归大规模语言模型(Large Language Models, LLM)，LLM 包含了多种自然语言处理任务中使用的深度学习模型结构，这些模型通过构建深层神经网络，利用现有的海量文本数据实现对文本的自动生成。GPT 的全称是 Generative Pre-Trained Transformer，其中自回归转换器(Transformer)架构是一种基于自注意力机制的神经网络模型，它通过在大规模无标注语料数据上进行“自监督(semi-supervised)”预训练，使模型具备基础的语言理解和生成能力[4]。

(二) 生成式人工智能的技术展望

人类社会的技术变革带来的主要是物质生产形式的进步，而以 ChatGPT 为代表的生成式人工智能直接模拟人类思维和思考方式，产生出想法，甚至对人类的思维习惯与行为模式产生直接的影响，这给人类社会带来了前所未有的冲击。随着技术的进步和创新，我们可以展望其在以下几个方向取得重要技术进展。

(1) 文本生成更智能化

相对于传统的聊天机器人，ChatGPT 更加智能化，具有高度拟人特征，具有高度的逻辑思维，甚至还具有一定的记忆力，可以持续地与用户进行交谈。与此同时，ChatGPT 由于其在处理自然语言等方面的优点，可以通过与用户的交互和反馈来对答案进行调整，从而持续改善其交流的准确性和对话体验。但在实际应用中也会以一种令人信服但完全编造的方式来生成文本，缺乏真正的理解能力，这就导致它对一些复杂的、需要深入理解的问题的难以解决，甚至会出现“一本正经的胡说八道”的现象，从而产生事实性错误[5]。此外，ChatGPT 所使用的词汇预测方法太过单一，本质上不具备对精确信息进行建模与推理的能力，尽管 ChatGPT 具有“举一反三”的自我学习功能，但由于其认知局限于语言模型所提供的大量文本语料，因此不可避免地会产生“认知”之外的问题。比如，ChatGPT 不能对超出语料库时限范围内的信息问答作出回应，也不能对还未发生的事情进行科学性预测[6]。为了克服上述困难，ChatGPT 还需要具备更透彻的理解能力和更精确的建模能力。例如，我们可以通过对数据进行清理来改善数据的学习质量，通过对模型参数的扩展来增强对模型的理解能力。在技术不断更新的过程中，ChatGPT 对于人类语言规则的破译和理解能力还在不断地提升，所生成的文本内容也会越来越智能化。

(2) 多模态快速发展

随着技术的不断发展，生成式人工智能正在从单一的语言生成逐渐转向多模态、具身化。多模态 AI 是指能够处理和理解多种类型信息的人工智能，如文本、图像、音频、视频等。这种 AI 不仅能够对单一数据类型的任务进行处理，还可以通过不同数据类型之间的关联与融合，实现一个综合、全面的理解多模态。AI 能够对各种不同类型的数据进行关联分析，可用于解决复杂问题。在多模态技术发展之初，就有不同模态的 AI 相互融合，比如图像识别和自然语言处理技术的结合。除了 ImageBind 打通六种模态之外，多数仍在探索文本与图像的融合，而且进展飞快。随着 OpenAI 于 2024 年 2 月推出了全新的视频生成大模型 Sora，这一推出标志着 AI 能力的重大飞跃，其技术复杂度的显著增加标志着包括文生文(T2T)、文生音频(T2M)、文生图(T2I)、文生视频(T2V)在内的多模态内容生成技术日趋成熟[7]。随着多模态技术的持续进步，多模态 AI 能够实现对各种类型数据的更为深度和有效的整合。该算法及模型具备强大的跨域适应性，能够有效地衔接并整合多种数据类型，进一步提炼出丰富的多模态共通特征。这赋予了 AI 深入解析和有效解决复杂问题的能力。多模态技术在未来在诸多创新领域的应用，必将开辟新的应用前景。然而，多模态 AI 的发展也存在一些问题。如何对不同来源的信息进行有效地整合与解释，是当前亟需解决的技术难题。此外，多模态 AI 模型所需的计算资源和数据量巨大，这也制约了其在实际应用中的普及和推广。

3. 生成式人工智能法律应用风险

生成式人工智能的兴起为人工智能的发展提供了全新的视角，它展现出空前强大的能力，但同时也带来一些新的法律风险，必须由法律对其进行回应和解决，只有在正视这些风险的基础上，才能维护法律的稳定性，实现国家治理与社会治理的价值。结合生成式人工智能目前暴露出的法律风险，本文选定三类常见法律风险予以具体分析。

(一) 数据安全问题

生成式人工智能是建立在大型的语言模型基础上的，它需要海量数据来学习和完善。以生成式人工智能的代表 ChatGPT 为例，其运作原理主要分为三个步骤：一是预先训练阶段；二是模型构建和微调阶段；三是输出处理、回答生成、输出处理阶段，在这三个阶段中都存在数据的收集、整合和分析，数据贯穿着生成式人工智能服务的整个动态过程，同时也伴随着种种与数据相关的法律风险。

(1) 预先训练阶段的数据来源合法性问题

虽然 ChatGPT 未公布训练数据集的具体信息，但我们仍可以推断出其训练数据集的主要来源。有研究者指出，OpenAI 主要基于维基百科、书籍、期刊、Reddit 链接、抓取大型数据集以及由 GitHub 等代码数据集、StackExchange 等对话论坛以及视频字幕数据集展开大模型训练[8]。ChatGPT 的研发公司 OpenAI 宣称，维护该系统运作的数据文本都来自公开数据，因此不会对个人、企业和国家数据安全造成威胁，但是作为一款对个人用户开放、收费的商业产品，ChatGPT 的模型数据收集来源是否被完全授权应用于商业活动仍然存疑[9]。随着社交媒体平台越来越隐私化，ChatGPT 等生成式人工智能将个人不愿意公开或仅对少部分人共享的个人信息、个人隐私、商业隐私以及肖像、名誉、作品等内容进行了预先训练，并使之成为可供公众使用的商业产品，其合法性还有待商榷。

(2) 模型构建和微调阶段的数据泄露风险

生成式人工智能文本生成依赖于自然语言模型，而自然语言模型在本质上取决于算法选择以及用于模型训练的海量数据库，其中包括了大量网民用户自主输入的海量数据。基于此，生成式人工智能的深度学习机制是以人类提供的反馈数据为基础，一旦用户在人机交互中输入隐私信息时，ChatGPT 就可能会自动将其导入自身的语料库中。当其他用户也检索相似信息时，就有可能造成个人的隐私内容被泄漏。尽管 OpenAI 公司已经做出了删除用户个人身份信息的承诺，但其具体是如何被删除的并不清楚，也无法对其进行验证和评价。目前，ChatGPT 还没有形成一套科学、高效的数据利用机制，难以确保语料来源的真实性，因此，在全球化、开放的信息传播环境下，极易造成数据泄露。

(3) 回答生成与输出处理阶段的数据伪造

虽然生成式人工智能的生成水平已经得到了明显的提升，但它在生成完全准确、真实和有效信息方面的能力仍具有很大问题。生成式人工智能为用户提供的最终信息就是在经过海量信息筛选后得出的一种简单、标准化的内容。所以，生成式人工智能不能对所产生的内容自身的真实性和正确性作出评判。该特性使得生成式人工智能能够制造出大量的错误信息，并将其用于恶意内容的生产，给人类的思考与行为带来极大的误导与消极影响。除此之外，通过生成式人工智能编造虚假信息而谋取私利的行为也经常发生。例如，由生成式人工智能生成的“换脸视频”将受害者的面部替换成其他人或物进行诽谤、威胁或勒索嫌疑人便是利用了 ChatGPT 编造虚假信息[10]。

(二) 知识产权问题

生成式人工智能在文本撰写、图像设计与音频制造等领域中表现突出，然而其广泛应用对现行的知识产权法律体系带来了新的挑战，其中最具有冲击的是，它在生成阶段对知识产权领域所构成的挑战。因为生成式人工智能本身就具有高度的智能化，因此在运算过程中，与之前的人工智能系统相比，其知

识产权的归属发生了颠覆性的变化。ChatGPT 是一种生成比分析式人工智能更强大的生成式人工智能，其内容生成过程主要包括内容自动化编纂、智能化修整加工、多模态转换、创意生成等环节，对出版的内容生产模式和内容供应模式产生直接影响。尽管 ChatGPT 的创造者中包含一些自然人的创作因素，在一定程度上满足了作品的构成要求，但是对于该类型的生成式人工智能所创造的作品能否被赋权，目前尚有争议，并且也缺乏具体的赋权认定标准研究。因此，生成式人工智能不可避免地面临着知识产权问题。

随着 ChatGPT 类生成式人工智能技术的不断升级，由此引发的知识产权争议问题还将继续拓展与深化，这不仅关系到 ChatGPT 类生成式人工智能文本生成模式的技术认定问题，也关系到人工智能是否可以作为法律主体的争论。ChatGPT 类生成式人工智能一旦取得了法律主体地位，需要对其知识产权保护体系进行全方位的调整，就当前科技发展的态势来看，ChatGPT 类生成式人工智能技术距离强人工智能阶段还有很长一段路要走，还不具备与人类思维相匹配的意识与能动性，所以无法撼动人类的法律主体地位，但是，随着人工智能技术的迅速发展，当生成式人工智能具备了“创作”的能力，传统知识产权的保护体系可能要进行全面调整。

(三) 虚假信息问题

在生成式人工智能视域下，虚假信息具有独特性。生成式人工智能所产生的虚假信息较以往更难治理，并且由于人工智能的深度伪造技术能使虚假信息极为逼真[1]。作为我国首个 AIGC 监管文件，《暂行办法》也将“不得生成虚假有害信息,不得侵害他人名誉权、荣誉权等”明确列入第 4 条，反映出生成式人工智能对于网络虚假信息监管存在潜在的威胁[11]。

生成式人工智能模型的运行需要海量的预训练数据作为学习基础，正如 GPT 训练数据集包含 1.17 亿个参数，而最新发布的 GPT-4 的单词回复量更是多达 25,000 个。随着参数量、单词回复量的飞速增长，开发人员已经无法对训练数据的真实性、客观性和合法性进行保证，而基于真伪难分数据基础上的人工智能生成虚假信息将会以爆炸性的速度增长。生成式人工智能模型学习的一个关键特性就是将机器最初的输出内容调整并作为后续训练的数据源，如此反复，直至形成所需的输出结果，这种不断调整完善的模式被称为“基于人类反馈的强化学习”模式，通过人工给予正反馈和负反馈来调整生成的内容，以便在特定的语境下生成特定的内容。但这些带有人类价值和个人喜好的人工反馈会加剧生成虚假信息、歧视性和带有偏见的信息。目前，此类问题在生成式人工智能的应用中已经显露出来。当生成的内容中混入了虚假信息，这些信息一旦被发布或被用户采纳，就会给社会秩序和公民合法权益造成严重的风险和危害。

4. 防范风险的法律规制路径

(一) 加强数据治理监管

《暂行办法》在第 7 条列举生成式人工智能服务提供者在进行模型训练时应当遵守的基本要求，具体包括三个方面即“不得侵害他人依法享有的知识产权”“涉及个人信息的，应当取得个人同意或者符合法律、行政法规规定的其他情形”“采取有效措施提高训练数据质量，增强训练数据的真实性、准确性、客观性、多样性”。具体来说，生成式人工智能收集和使用个人数据是以取得个体独立同意的情况下。目前，对于生成式服务提供者而言，完全遵循“知情–同意”的原则并不切合实际，可采取强制性的录音义务、强化保密措施、泄漏后的快速处理措施等措施来加强对其的监督。另外，还可以通过注重保护个人资料的权利，来预防生成式服务提供者对个人资料的滥用。

针对生成式服务提供者。首先,生成式服务提供者在获取第三方平台(如网站、应用等)上的个人资料时，应保存阅读记录及与第三方签署的数据提供服务协议，并备案。其次，生成式服务提供者在数据流

通、处理等环节，要采取有效的防范措施，应以必要的手段防止数据泄漏。最后，生成式服务提供者在发现数据泄漏的情况下，应及时通知数据监管机构，告知数据当事人，并修复相关漏洞，尽可能将影响降到最低。

针对数据主体。首先，赋予生成式服务使用者数据访问权，给予其必要数据访问的合法渠道。根据我国个人信息保护法规定，个人有权向个人信息处理者查阅、复制其个人信息，因此，使用者应该享有数据访问权，以知悉其个人数据的使用情况。其次，生成式服务提供者间接获取的数据存在陈旧的、不准确的问题，应该赋予使用者更正和补充个人信息的权利，以避免数据被误用。最后，在数据泄漏等突发情况下，生成式服务使用者享有知情权，即使用者有权知晓数据被窃取，以及被何人窃取等讯息。

(二) 完善知识产权的相关法律法规

生成式人工智能的应用对我国知识产权保护体系产生了很大影响。就目前的科技发展状况而言，虽然它并不能完全推翻知识产权保护体系，但是，对于已出现的知识产权争端，仍有必要从法律层面作出迅速的反应与规制。

第一，适应生成式人工智能技术的知识产权法规。为了应对生成式人工智能面临的法律风险，必须对其进行相应的规制，以适应该技术的发展和应用。为此，首先可以修订著作权法，针对生成式人工智能生成的知识产品，明确著作权的归属和保护范围。不仅要肯定生成式人工智能的创作者的身份，使其作品在《著作法权》中得到保护，又要将生成式人工智能作为著作权原始主体，换言之，除特殊情况外，生成式人工智能作为作品创作者，成为著作权的原始主体。相应的权利、责任都由生成式人工智能来承担[12]。其次，针对生成式人工智能生成的相关知识产品，加强知识产权的保护力度。最后，针对生成式人工智能生成的技术创新，适应其特点和应用方式，优化专利制度，并明确专利的申请和保护程序。

当然，无论是从著作权还是行业秩序的角度来看，对人工智能生成内容进行保护都有其必要性，但同时也要注意纳入保护范围的门槛要求，以免造成司法资源的浪费。

第二，知识产品创作责任和权益的明确。生成式人工智能的知识产品创作涉及多方利益相关者，需要明确各方的责任和权益。具体而言，第一步就应当制定责任规范，即对生成式人工智能的开发者、使用者、平台运营商等各方的责任和义务进行界定，明确他们在创作和使用知识产品过程中的法律责任。其次，应当建立合规的权益分配机制，制定基于生成式人工智能产生的知识产品的权益分配机制，明晰各方在知识产权和经济利益等方面的权益。通过合同、协议、平台规则等方式，对知识产品的归属和权益分配进行明确。最后，应站在相关平台的立场上，强化对生成式人工智能服务的平台的管理和监管，以确保用户的合规操作。平台应当制定相关规则和审核机制，加强对知识产品的生成和传播的监督和控制，防范违法违规行为。

(三) 构建完善的法律责任体系

在解决生成式人工智能生成虚假信息的法律问题时，首先要明确这些虚假信息的生成并非人为因素所导致，而是源于技术模型的运行本质。《暂行办法》强调了增强生成式人工智能服务的透明度，并致力于提升生成内容在准确性和可靠性方面的标准。尽管 ChatGPT 凭借其庞大的数据存储，能提供丰富的信息，但当前的技术局限性制约了其识别和验证所有信息真实性的能力。鉴于当下的法律框架和生成式人工智能的技术现状，赋予生成式人工智能独立承担虚假信息法律责任并不适宜，鉴于其本质上并未具备法律主体地位。应对虚假信息的扩散，需要技术手段与多元主体的协同作用。从技术实施层面出发，首要任务是在现有技术水平的限制下，极力抑制和排除伪造信息，通过优化内容生成算法，极力降低虚假信息的存在率，并持续推动技术革新。这应被视为生成式人工智能技术未来发展的一个核心导向。其次，程序开发者、运营商和服务提供商应履行责任，对信息的真实性实施预警，促使用户意识到人工智能生成内容的真实性并进行必要的验证，以防止产生误导。

在生成式人工智能的监管框架中，有效管理虚假信息的法律隐患关键在于构建完善的法律责任体系，明确各个责任主体的职责，特别是要严谨地定义平台作为“守门人”的法定责任[13]。责任的确立首先要求明确责任主体，这涵盖了网络服务供应商、网络技术供应者及技术使用者[14]。针对虚假信息的管控，网络服务提供商应采取多维度的预防策略：首先，强化用户行为教育，强调其法律责任，引导用户践行合规使用原则；其次，实施严格的实名认证制度，以确保在发生侵权事件时能追踪责任源头，形成威慑效果；最后，优化用户账号管理，严查并剔除疑似机器人账户，迅速处理和封禁违规账号，防止其重新注册。

5. 结语

总之，生成式人工智能将继续迅猛发展，法律风险问题也将持续存在，这对人工智能的监管提出一系列挑战，还需进一步深化研究。本文聚焦生成式人工智能的技术发展和应用风险，提出防范风险的法律规制路径，未来还需要从我国国情和国际规则出发，及时调整和优化相关法律框架和标准，以适应新兴问题和技术挑战，构建生成式人工智能的本土化法制体系。

参考文献

[1] 张莉莉, 华志宇. 生成式人工智能的法律风险及其规制[J]. 行政与法, 2024(4): 88-100.

[2] 董扣艳, 张雨晴. 生成式人工智能发展与治理的哲学省思[J]. 福建师范大学学报(哲学社会科学版), 2023(4): 48-63.

[3] 宋华健. 论生成式人工智能的法律风险与治理路径[J]. 北京理工大学学报(社会科学版), 2024, 26(3): 134-143.

[4] 李耕, 王梓烁, 何相腾, 等. 从 ChatGPT 到多模态大模型: 现状与未来[J]. 中国科学基金, 2023, 37(5): 724-734.

[5] 张熙, 杨小汕, 徐常胜. ChatGPT 及生成式人工智能现状及未来发展方向[J]. 中国科学基金, 2023, 37(5): 743-750.

[6] 董扣艳. 生成式人工智能对网络意识形态安全的挑战及应对[J]. 实事求是, 2024(1): 79-85.

[7] 邹开元, 沈阳. Sora 的技术复杂度突破: 人工智能生成内容的新质生产力转型[J/OL]. 西南交通大学学报(社会科学版), 1-13. https://kns.cnki.net/kcms/detail/51.1586.C.20240422.1057.002.html, 2024-05-11.

[8] 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.

[9] 段广远, 刘伟丽. 生成式人工智能领域中的数据安全风险及刑法保护[J]. 中共郑州市委党校学报, 2024(1): 62-67.

[10] 李杨. 生成式人工智能风险的法律类型化治理[J]. 延安大学学报(社会科学版), 2024, 46(1): 31-38, 57.

[11] 朱嘉珺. 生成式人工智能虚假有害信息规制的挑战与应对——以 ChatGPT 的应用为引[J]. 比较法研究, 2023(5): 34-54.

[12] 何萍, 刘继琨. 法域协调视角下生成式人工智能产物涉知识产权问题研究——以 ChatGPT 为例[J]. 法治研究, 2024(3): 47-56.

[13] 舒洪水, 彭鹏. ChatGPT 场景下虚假信息的法律风险与对策[J]. 新疆师范大学学报(哲学社会科学版), 2023, 44(5): 124-129.

[14] 鲍家琴. 生成式人工智能视域下虚假信息的法律风险及规制[J]. 中国价格监管与反垄断, 2024(3): 90-93.