Dispute Settlement 争议解决, 2025, 11(8), 21-27
Published Online August 2025 in Hans. https://www.hanspub.org/journal/ds
https://doi.org/10.12677/ds.2025.118239

Hans汉斯

生成式人工智能应用的社会风险与法律因应

梁慧仪，庄文锐，吴冀瑜

佛山大学法学院，广东 佛山

收稿日期：2025年6月26日；录用日期：2025年7月26日；发布日期：2025年8月6日

摘　要

生成式人工智能通过深度学习和大数据训练，能够生成高质量的内容，但其应用引发在个人隐私、算法偏见、间接侵权、权属界定等多方面引发了社会风险。本文从生成式人工智能的技术原理着手，将生成过程拆分为数据获取、数据处理、成果生成与成果存储四个阶段；各阶段社会风险集中体现为个人隐私风险、算法偏见风险、间接侵权风险与生成成果的权属界定风险。通过进一步探讨了这些风险的成因，本文提出要明确责任主体、健全多元监管机制和完善法律保障体系等规制进路，以期在保障个体权益和公共利益的同时，促进人工智能产业的良性发展。

关键词

生成式人工智能，社会风险，法律规制

Social Risks and Legal Responses to Generative Artificial Intelligence Applications

Huiyi Liang, Wenrui Zhuang, Jiyu Wu

School of Law, Foshan University, Foshan Guangdong

Received: Jun. 26th, 2025; accepted: Jul. 26th, 2025; published: Aug. 6th, 2025

Abstract

Generative artificial intelligence can generate high-quality content through deep learning and big data training, but its application triggers social risks in terms of personal privacy, algorithmic bias, indirect infringement, and ownership definition. In this paper, starting from the technical principle of generative AI, the generation process is divided into four stages: data acquisition, data processing, result generation and result storage; the social risks in each stage are centered on the risks of personal privacy,

文章引用: 梁慧仪, 庄文锐, 吴冀瑜. 生成式人工智能应用的社会风险与法律因应[J]. 争议解决, 2025, 11(8): 21-27.
DOI: 10.12677/ds.2025.118239

algorithmic bias, indirect infringement, and ownership definition of the generated results. By further exploring the causes of these risks, this paper proposes to clarify the main body of responsibility, improve the multifaceted regulatory mechanism and perfect the legal protection system and other regulatory approaches, with a view to promoting the benign development of the AI industry while safeguarding the rights and interests of individuals and the public interest.

Keywords

Generative AI, Social Risk, Legal Regulation

1. 引言

人工智能(Generative AI)技术迅猛发展，并广泛应用于图像生成、文本创作、语言翻译、音乐创作以及虚拟角色建模等多个领域。随着 GPT 系列、DALL·E 和 DeepSeek 等模型的商用化，人工智能在生产力、娱乐业、教育和医疗等领域的潜力被深入发掘。生成式人工智能通过深度学习和大数据训练，实现了对输入信息的理解、处理和再创作等活动，正以人类难以企及的速度和精度生成内容。

然而，随着生成式人工智能的广泛应用，一系列法律问题逐渐凸显。例如，在版权问题上，AI 生成的作品应如何界定权利归属？在隐私保护层面，AI 生成的内容是否可能暴露敏感信息？更为复杂的是，当生成的内容导致虚假信息传播或其他危害时，如何追责[1]？这些问题不仅挑战现有的法律体系，也迫使政策制定者和行业监管者重新审视技术带来的法律空白。据此，本文将基于对生成式人工智能的技术理论的分析，探讨其在应用过程中可能面临的法律风险及风险成因[2]，进而提出相应的法律机制予以纾解。

2. 生成式人工智能应用的技术背景

生成式人工智能的工作原理可以概括为四个主要阶段：数据获取、数据处理、成果生成和成果存储。每个阶段涉及不同的技术手段和算法设计，但它们共同的基础是利用神经网络模型对大规模数据进行训练，以生成高质量的、具有实际应用价值的内容。以下将逐步介绍各阶段的具体流程，并通过流程图展示其运作机制。

2.1. 数据获取

生成式模型的效果高度依赖于训练数据的质量与覆盖面。当前数据来源主要包含两类：一是互联网公开数据资源；二是经过专业整理的机构数据，包括科研数据库、企业专有数据及政府开放数据等。

在数据获取阶段需重点解决主要会进行以下流程：首先，针对网络公开数据，采用分布式爬虫技术实现多源异构数据的自动化采集，在采集过程中涉及数据版权合规性问题；其次，结构化数据调用通常通过 API 等标准化接口实现；最后，原始数据必须经过严格的清洗流程，包括去重、异常值处理和数据标注等环节，确保模型训练效果[3]。

2.2. 数据处理

在数据获取完成后，人工智能模型需要对数据进行预处理，以便在训练过程中有效学习。数据处理

的核心目的是将原始数据转换为模型能够识别和理解的结构化输入形式。数据预处理是对原始数据进行规范化和标准化处理，删除冗余信息，填补缺失数据。对于防止模型过拟合或在特定数据上表现不佳，则通常会采取数据增强操作。对于图像数据，可以进行旋转、翻转、缩放等操作；对于文本数据，可能会使用同义词替换或随机插入词语的方式扩充数据集。此外，对于人工智能模型无法直接处理原始数据，需要通过特征提取将数据转换为模型可接受的特征向量。

2.3. 成果生成

数据处理完毕后，生成式人工智能模型开始训练，并基于输入生成相应的成果。这一阶段的核心是模型的训练和推理过程。当前，最为流行的生成模型架构是生成对抗网络(GAN)和变分自编码器(VAE)等。

关于模型训练，生成式模型会在训练阶段对输入数据进行大量的学习迭代。以生成对抗网络为例，训练过程中会涉及两个网络的对抗：生成器(Generator)生成数据，判别器(Discriminator)判断生成的数据是否真实。通过这种对抗训练，生成器最终可以产生与真实数据相似的高质量成果。基于此训练完成，生成式模型可以接受新的输入数据，并基于输入信息生成成果。例如，GPT 模型可以根据输入的文本提示生成具有连贯语法和语义的文章；DALL·E 模型则可以根据文本描述生成相应的图像。

2.4. 成果存储

生成的成果需要妥善存储和管理，以便后续使用或进一步处理(如图 1)。对于生成式人工智能的应用而言，成果的存储不仅是数据的简单保存，还涉及到数据的版本控制、结果的可追溯性以及隐私保护。

生成的文本、图像等成果通常存储在数据库中，便于检索和后续的分析。其次，关于版本控制与追溯，尤其对于企业的应用上，生成式 AI 的成果可能在多个版本中产生，因此需要通过版本控制工具追溯每一个生成过程，确保结果的可验证性。再次，在生成模型使用过程中，可能会涉及到用户的个人数据，因此需要对生成成果中的敏感信息进行加密处理，防止未经授权的访问或数据泄露[4]。

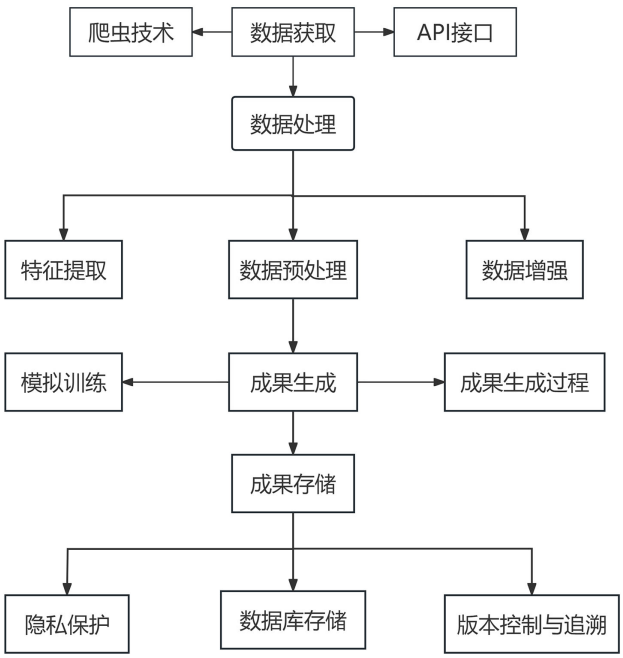

Figure 1. Storage and management of generative AI results
图 1. 生成式人工智能成果的存储与管理

3. 社会风险的表征与表现

3.1. 个人隐私风险

生成式 AI 的迭代发展离不开海量的、高质量的数据“喂养”，需要来源于公共互联网和人工标注的数据资源作为训练数据，而这些数据隐含着诸多不确定的法律风险，直接或间接地威胁着个体隐私、个人信息权益等方面的安全。正是由于生成式人工智能所需数据库十分庞大，广泛覆盖了各个方面，因此无孔不入的数据来源可能会导致个人隐私被广泛地泄露，既有可能是个人隐私信息垂直泄露，也有可能是生成式人工智能研发者通过购买某些特定领域的数据库而导致个人隐私信息被场景化泄露。2023 年 3 月 20 日，ChatGPT 因 Redis 客户端开源库错误发生数据泄露事件，使部分用户不仅能够看到其他用户的聊天记录，还能够看到电子邮件地址、支付地址、信用卡最后四位等与支付相关的信息。可见，在有些生成式人工智能数据库中，模型研发者对公开个人信息的利用往往超出了最初的场景脉络与用户期待，甚至超过了法律的底线。

3.2. 算法偏见风险

“信息茧房”是推荐算法广泛应用之下难以规避的结果。通过对人类价值判断与意识形态的渗透，生成式人工智能将经过处理和筛选的内容与人们的价值选择、意识形态相勾连，使人们在无意识时接受人工智能的潜在规训[5]。从内容生成准备阶段的价值预设与文化选择，到内容生成的反馈与应用阶段都渗透了掌握话语权力者以及相关利益者的喜好。尤其是心智尚未成熟的未成年人，难以对其接触的信息进行甄别和筛选，无法留下可靠、真实信息并排除对其不利的虚假有害内容。长此以往人类将失去对其接触信息是否真实的判断力，进而失去对社会最基本的信任。

3.3. 间接侵权风险

间接侵权风险具有隐蔽性。在生成式人工智能的个人信息保护中，除了还存留着传统人工智能个人信息“告知形同虚设”的风险外，用户面临的告知同意风险也许更加严峻。虽然生成式人工智能在收集数据时也许尽到了“告知同意”的义务，但是其隐蔽的处理规则、目的、方式、范围，让个人信息主体很难判断他们的信息将面临什么处理活动。在生成式人工智能中，个人信息是否被处理以及通过，何种方式处理，通常只能通过回应某些提示时生成的特定输出才能显现[6]，这意味着在没有进行深入调查的情况下，确定个人信息是否被违法处理具有挑战性。从而造成了用户个人隐私通过算法处理后被过度攫取或者超出了用户信息用途授权范围的风险。

3.4. 权属界定风险

生成式人工智能技术所生成的内容也面临着权利归属问题。人工智能究竟能不能作为知识产权的主体享有知识产权，承担相应的责任，学界仍众说纷纭。在实践过程中，也出现理论与实际相悖的现象。OpenAI 公司和微软公司所创建的人工智能代码辅助工具 Copilot，因代码使用问题而引发了纠纷。根据欧盟和美国法律规定，人工智能不能作为知识产权主体，最多只能算是创作辅助工具[7]，法律拒绝承认其作者身份，但是在实践中生成式人工智能技术却屡屡突破法律规定。例如，在学术论文投稿中，一些学者将生成式人工智能技术列为并列作者，ChatGPT 甚至出现在多篇《Nature》杂志作者栏中。对于生成式人工智能生成的内容是否构成原创作品、其版权归属如何确定等问题，法律尚无明确规定。

4. 生成式人工智能的风险成因

4.1. 数据收集和处理阶段

一方面，基于生成式人工智能的原理，生成式人工智能需要海量数据信息，然而在收集数据的过程

中，获取数据一方与被获取数据一方之间的联系并不密切，用户对于自己的数据被运用于 AI 训练存在不知情的情况，缺乏一定的知情权基础。

生成式人工智能在收集信息的过程中，如果不加以干预，有可能造成信息滥用，进一步侵犯隐私权。生成式人工智能造成个人隐私风险主要是来源于两个领域：(1) 收集、处理数据时来自不同数据源导致了风险的产生，人工智能数据主要来源于公民个人信息的汇集(人脸、指纹等)、开放的公共数据、开发者通过爬行手段自行获取等途径，尽管在我国的《个人信息保护法》中规定，个人信息的处理者需要经得被获取人同意才可以进行处理，但在现实的应用当中，收集个人信息的边界具有模糊性，使得这一规定难以落实。(2) 生成式人工智能的处理分析可能涉及个人隐私，在算法模型层面，一些原本看起来安全的个人信息，但经过模型的推理演算，也可能有个人隐私风险的隐患[8]。

另一方面，运用于训练文本生成模型的数据缺乏多样性，容易导致算法偏见与算法歧视。在生成式人工智能的运作过程中，算法正渗透到我们的社会生活当中，算法并非客观的。作为一种数学语言的表达，算法的偏见主要来源于算法设计本身、算法依据的数据和机器自主学习三个方面。人为预算的方式使算法进行运算，算法的开发者极可能将自身持有的偏见嵌入算法当中，而算法所依据的社会风气、制度体系、文化差异等信息，基于数据集的代表性不同会引发系统性偏差，导致不同特征的群体内容输入具有显著差异，而机器学习的过程又进一步强化偏见，一系列运行过程从而最终影响生成内容的可信度。机器学习的过程并非是一个中立的信息处理过程，相反，它会进一步强化已有的偏见。在机器学习不断迭代优化的过程中，带有偏差的数据会持续影响模型的参数调整，使模型对不同群体的认知和处理方式逐渐偏离客观公正的轨道[9]。

4.2. 成果存储和输出阶段

人工智能具有效率极高、产量极高等特点，目前已经在文学艺术等领域生产了大量成果，然而关于人工智能生成物的著作权争议问题日趋激烈，对于人工智能生成物(AIGC)的权属界定众说纷纭。

2019 年，腾讯 Dreamwriter 案，上海盈讯公司在其运营的网站传播由腾讯智能写作软件 Dreamwirter 生成的财经报道，引发人工智能生成稿件著作权纠纷；菲林诉百度案，2018 年 9 月，原告在网络平台上公布《影视娱乐行业司法大数据分析报告电影卷？北京篇》，被告未经允许在互联网上发布了涉案文章，该涉案文章由原告收集数据借助智能软件自动生成而成；AI 纹身图案是人工智能生成图片著作权侵权第一案，原告将利用 Stable Diffusion 模型制作的图片命名为“春风送来了温柔”发送到社交软件小红书，2023 年 3 月，被告擅自将该图片运用到发布的文章《三月的爱情，在桃花里》并删除水印，引发了著作权纠纷[10]。这三个案件的争议焦点主要是围绕着人工智能生成物是否享有传统意义上的著作权以及著作权应归属于谁这两大问题产生。

除此之外，生成式人工智能所生成的作品有可能与现有受著作权保护作品构成相同或相似，若未经版权所有者授权或未遵循合理使用原则，有可能会引发侵权风险。

5. 规制进路

5.1. 明确责任主体，推动协调共治

在在人工智能领域所运行的过程中，需要明确包括算法设计者、提供者、使用者在内的各方责任主体的法律责任[11]。其一，企业作为人工智能产品的主要供应方，在技术研发与市场运营中占据主导地位，应当承担起关键责任。在产品与服务上，确保技术架构稳定可靠，并防止数据泄露、算法偏差等问题，满足安全性与可靠性要求。其二，个人使用者在使用人工智能产品与服务时，必须严格遵守相关法律法规，避免利用其从事非法信息传播、网络诈骗、侵犯隐私等违法活动，确保个体层面责任可追溯，防止

个人不当行为扰乱人工智能领域秩序、破坏法律规范。其三，政府在人工智能规制中起着极为重要的作用，通过制定系统且适宜的政策法规，明确各方行为准则，为人工智能产业营造良好发展环境。各方责任主体协同合作，协调利益冲突，构建多元主体协作的规制格局，推动人工智能领域健康、合法且持续地发展，为社会经济发展注入稳定动力与有力支撑。

5.2. 健全多元监管机制，实习全链条合法性监管

随着人工智能技术以迅猛之势不断向前发展演进，近年来，我国已逐步开启并积极深入地探索人工智能领域风险监管治理的有效路径与模式。一方面，从立法层面考量，可以深入研究并考虑通过制定专门的人工智能法律法规来确立其基本规则和标准体系，从而为数据保护、算法透明度、知识产权保护等多方面的监管工作提供坚实且明确的法律依据与制度支撑。在数据保护方面，立法应详细规定数据收集、存储、使用以及共享的合法流程与边界条件，明确数据主体的权利。

另一方面，在监管机构设置与职能强化方面，可以考虑建立或进一步强化专门的人工智能监管机构，赋予其明确且独立的监管权力与职责范围[12]，专门负责监督人工智能技术的研发进展、应用场景以及市场运营等各个环节，确保其始终符合法律法规的要求以及社会伦理道德规范[13]。对人工智能产品的研发过程进行审核与评估，检查其是否遵循了数据安全与隐私保护的相关规定，算法设计是否存在歧视性或不公平性的因素；在产品应用阶段，对其应用场景进行合规性审查，防止人工智能技术被滥用或用于非法目的。

5.3. 构建“分级分类”风险治理体系，优化风险预防机制

《生成式人工智能管理暂行办法》的出台，其中确立的包容审慎和分类分级监管原则，清晰且明确地表明了我国对于生成式人工智能的基本态度与监管导向。为切实将这些原则性规定落到实处并转化为具体有效的监管行动，充分考虑到人工智能风险复杂多样且不断变化的显著特点，可以积极借鉴欧盟在人工智能监管方面的先进经验与成熟做法，对人工智能进行更为细致且具体的分类划分。依据人工智能技术的不同应用领域以及所蕴含的风险等级差异，实施分类分级监管策略，精准地划分出低、中、高风险区间，并针对不同风险区间制定差异化的监管措施与要求，以此确保监管工作的有效性和针对性得以充分体现[14]。只有在企业满足一系列严格的技术、安全、伦理等多方面标准与条件，并获得监管机构的专门许可后，方可进行研发、生产与应用，且在整个生命周期内都要接受监管机构的持续监督与动态评估[15]。

参考文献

[1] 匡恺, 刘力铭. 生成式人工智能的风险特征及语义图像: 基于网络社区讨论的计算文本分析[J]. 河北大学学报(哲学社会科学版), 2024, 49(6): 147-160.

[2] 苏文成, 洪舒悦, 卢章平, 等. 人工智能风险体系与模块化评价指标构建实证研究[J]. 情报杂志, 2025, 44(1): 136-145+154.

[3] 王海洋. 生成式 AI 训练数据的法律风险及其元规制[J]. 浙江社会科学, 2024(9): 50-63+157-158.

[4] 靳梦戈. 去中心化背景下数字身份识别的法律风险及其应对[J]. 四川师范大学学报(社会科学版), 2024, 51(5): 69-79+202.

[5] 代金平, 覃杨杨. ChatGPT 等生成式人工智能的意识形态风险及其应对[J]. 重庆大学学报(社会科学版), 2023, 29(5): 101-110.

[6] 张涛. 生成式人工智能中个人信息保护风险的类型化与合作规制[J]. 行政法学研究, 2024(6): 47-59.

[7] 徐磊. 生成式人工智能技术的多元风险及其协同治理研究[J]. 图书馆建设, 2025(2): 59-69.

[8] 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.

[9] 刘友华. 算法偏见及其规制路径研究[J]. 法学杂志, 2019, 40(6): 55-66.

[10] 刁胜先, 秦兴翰. 论人工智能生成数据法律保护的多元分层模式——兼评“菲林案”与“Dreamwriter 案” [J]. 重庆邮电大学学报(社会科学版), 2021, 33(3): 41-53.

[11] 毕文轩. 生成式人工智能的风险规制困境及其化解: 以 ChatGPT 的规制为视角[J]. 比较法研究, 2023(3): 155-172.

[12] 张欣. 生成式人工智能的算法治理挑战与治理型监管[J]. 现代法学, 2023, 45(3): 108-123.

[13] 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.

[14] 郭小东. 生成式人工智能的风险及其包容性法律治理[J]. 北京理工大学学报(社会科学版), 2023, 25(6): 93-105+117.

[15] 程乐. 生成式人工智能治理的态势、挑战与展望[J]. 人民论坛, 2024(2): 76-81.