生成式人工智能的数据风险及法规研究

李新颖 *
杭州师范大学　浙江杭州　310000

摘　要：随着生成式人工智能技术的颠覆性发展，使得产业结构得以优化升级，但是在人工智能技术应用的各阶段也带来了诸多数据安全风险。国内国际都出现了对于人工智能风险规制等方面的学术探讨，以便于应对生成式人工智能引发的数据安全、隐私保护等一系列的难题。为了应对上述风险难题，在现阶段，我国生成式人工智能治理应坚持风险防范与研发并行原则，遵循基本的伦理道德准则，针对生成式人工智能制定专门的法律文件、构建统一的监管体系、设立自治机制等手段丰富生成式人工智能的预防监管措施，以期能够更好的应对风险与挑战，推动生成式人工智能的协调健康发展。

关键词：生成式人工智能；风险；风险规制

1. 问题的提出

以 ChatGPT 为代表的生成式人工智能（Generative Artificial Intelligence），能够模拟和生成类似人类创造的内容，如图像、音频、文本等。[1] 这些生成式人工智能模型的核心是理解数据和任务的能力，它可以在与人们对话交流的过程中，不断地根据与人类交流的反馈，进行上下文互动。，有些生成式人工智能还可以完成视频剪辑等创造性的工作任务。生成式人工智能搜集大数据集，其内容生成能力也会持续提升。目前，以 ChatGPT 为代表的大语言模型在回答问题、提供建议、总结和优化文本等文本生成任务中已经达到或超过了人类水平。[2]

生成式人工智能一方面以其强大的互动性和智能生成性，加速与人类社会的融合。另一方面，存在许多诸如数据泄露、充斥虚假信息等缺点，滋生出许多的风险与问题。各国对法律与制度对生成式人工智能的回应逐渐从“替代”与“辅助”议题转向“风险”议题。[3] 在生成式人工智能的早期阶段，法学研究者着重探讨生成式人工智能能否替代人类，是否能够完美的完成人类交代的工作任务。在“风险”议题研究阶段，应当探讨的是法律在面对生成式人工智能背后出现的风险问题，如何预防规制风险，让人工智能更好的促进人类社会的发展而不是让替代甚至是威胁人类。

2. 生成式人工智能带来的风险与挑战

2.1 数据安全风险

首先，生成式人工智能在数据收集及标注的准备阶段存在风险。[4] 生成式人工智能运行的基本逻辑是人工智能使用者利用人工智能技术，通过机器学习和自然语言处理等算法，自动生成个性化内容。如果训练数据来源不均衡，标注数据可能会存在偏差，生成式人工智能算法可能会学习到人类偏差，最终损害生成式人工智能的准确性。[5] 例如，ChatGPT 的训练数据集 96% 来自英语文本，在处理指令过程中，可能造成意识形态潜在偏差，此外，生成式人工智能在运行和使用过程中可能会生成大量虚假信息。生成式人工智能可以根据用户自动生成个性化内容，成本更低、质量更高。但是受市场需求支配，目前的生成式人工智能大多是聊天式的人机交互模型，让用户以便捷的形式发送指令，生成相应信息，以便扩大用户规模，实现最大程度的盈利。这类信息往往以满足用户个性化需求为出发点，匹配度优先于准确度，可能会出现真伪难辨的情况，除此之外，利用生成式人工智能编造虚假信息谋取个人利益的案件也频频发生。[6]

2.2 隐私安全与合规性问题

生成式人工智能在数据收集阶段会侵犯个人隐私。生成式人工智能在前期的训练阶段可能会收集大量的数据信息，有些甚至是未经个人用户同意的信息，造成侵犯隐私的问题。例如，在广泛的信息传播平台，微信公众号上发布的信息鱼龙混杂，会存在运用生成式人工智能形成的信息，这些信息未经他人同意擅自使用，使得个人信息难以得到保护，严重使可能会造成侵权。生成式人工智能在交互使用过程中存在隐私泄露风险。[7] 以“文心一言”、“通义千问”为例，在初始阶段注册使用时，要求事前收集用户的个人信息，不同意难以进行应用使用，此外，在正式使用阶段，用户所输

入的个人信息也在不经意间被收集使用，以上的行为都不会明确的通知使用者。

3. 生成式人工智能的初步规范

生成式人工智能不仅带来了新的机遇与发展前景，但是也滋生出许多的风险和问题。国内外有关部门在监管层面都给予高度重视，并且为此采取一系列的措施。[8]

3.1 国外相关规定

对面对生成式人工智能带来的风险难题，欧盟积极回应并做出一系列的人工智能立法探索。欧盟委员会在 2021 年 4 月发布了《关于制定人工智能统一规则》立法提案，经过多方讨论，最终通过了通过了《人工智能法案（AI Act）》，设置了生成式人工智能专门规制条款，包括禁止面部识别、新的透明度要求等。[9]

美国商务部以发展引导为方向，侧重于对人工智能的开放性管理，让企业和市场对人工智能的发展进行先导，对于具体的风险控制，美国发布了《人工智能权利法案蓝图》，制定了《人工智能风险管理框架》，同年 4 月发布了“人工智能问责政策”征求意见稿，对于人工智能（ChatGPT 等）监管问题征求意见。

加拿大联邦首先对生成式人工智能进行意见收集，呼要求一些利益相关者就与生成式 AI 相关的版权问题提供技术证据。加拿大政府认为人类可以在人工智能技术的帮助下制作的作品中贡献足够的技能和判断力，从而被视为该作品的作者，有必要对人工智能生成物的版权问题进行进一步规定，进一步立法可以通过重新考虑如何定义作者，甚至不依赖作者身份，来澄清人工智能生成作品或人工智能辅助作品的优先所有权。

3.2 我国相关规定

近些年来，我国在发展新质生产力的要求下，积极进行法治研究工作，在人工智能领域主要是以纲领性的文件为主导，形成了多角度，全方面的综合治理体系，在生成式人工智能立法方面颇有发言权。[10]

一方面，我国先后颁布《关于加强互联网信息服务算法综合治理的指导意见》，明确了我国算法治理目标是建立治理机制健全、监管体系完善、算法生态规范的算法安全综合治理格局；《关于加强科技伦理治理的意见》提出伦理先行的治理要求；以及《新一代人工智能发展规划》《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》等促进人工智能应用发展的文件。另一方面，《中华人民共和国民法典》则是起到总领原则作用；《数据安全法》明确算法治理的原则，要求技术促进经济社会发展，符合社会公德和伦理；《个人信息保护法》要求对个人信息加大保护。加之其他部门联合出台的《互联网信息服务算法推荐管理规定》重点治理以“大数据杀熟”为代表的算法歧视性决策；《生成式人工智能服务管理暂行办法》关注生成式人工智能的规范性发展。[11] 特别指出《生成式人工智能服务管理暂行办法》是较为全面的规制。通过多方努力，生成式人工智能的法律体系也逐渐完善。

3.3 生成式人工智能法规的困境

虽然我国在近些年来发布了一系列人工智能领域的法律规范，在生成式人工智能法律规范方面较为领先，但总体上看，仍存在诸多制约与不足。

首先，生成式人工智能法律规范内容重叠。在《暂行办法》中，对于生成式人工智能提供这的法律责任没有分类分级管理，导致在不同数据处理情境下，AI 提供者的具体责任方式散见于《网络安全法》《个人信息保护法》等，《暂行办法》本身无法给出确切的指引。[12] 其次，《暂行办法》对生成式人工智能的产出没有具体的执行标准和法律责任规定。虽然在文件中规定了对于说呢过程呢过的图片和视频进行一定的标识，但是没有详细规定具体的执行标准，导致在现实中此类标识易于规避，适用性不强。此外，法律规制主体权责界限模糊。[13]《暂行办法》是由多个部门联合发布，这就导致各个部门都有监管职责。可能会出现各个部门竞相监管或各个部门推脱监管的现象，这种不确定的治理模式不利于生成式生成式人工智能的有效管控与监管工作的协调推进。最后，未形成完备的统一的责任承担体系。《暂行办法》第九条要求生成式人工智能提供者承担网络信息内容生产者的责任，与传统的智能产业领域服务提供者与内容生产者二元分立的情况大不相同，生成式人工智能在运行阶段面临着数据过度收集、非法收集、数据滥用等风险隐患，基于生成式人工智能运行过程的整体性，不能拆分规制，需要建立一个统一的生成式人工智能数据风险责任体系。况且生成式人工智能不具备对生成内容的完全控制，将给生成式人工智能分级管理及责任治理带来新的矛盾。

综上，我国目前的生成式人工智能法律内容不完整、监管主体职责模糊，规制手段混乱以及法律责任规则主体复

杂等问题，亟需在理论与实践方面进行深入研究。[14]

4. 生成式人工智能的法规路径

我国目前的生成式人工智能法律规范较为全面，但仍然难以应对具体现实化的实践问题，应继续探寻完善的风险防范路径。

4.1 完善人工智能领域的专门性法律。鉴于生成式人工智能类型多样且复杂多变，对生成式人工智能所导致的风险制定专门的法律，实现人工智能领域法治治理体系化的布局。[15] 对于一些原本文件中的参见条款进行删改，对于在法律文本中规范重复的予以少删除，在法律文本中规定不清晰的进行多方主体的明确，保证生成式人工智能得到准确性的应用。此外，在制定法律时应当对生成式人工智能进行一定的阐释，运用通俗易懂的语言让用户理解算法的基本逻辑，算法潜在的风险等，增加生成式人工智能的透明度，在合理的条件下，为用户创建申诉渠道，丰富投诉人的投诉举报方式，使用户能够及时维护自身的合法权益。

4.2 建立统一的生成式人工智能风险责任体系。运用分类方法合理界定责任主体，是管理生成式人工智能运行阶段风险的有效策略。在这一阶段，涉及服务提供者、技术研发人员、数据供应者及系统操作者等多个相关方。通过分类管理，明确各自的责任范围，能提升对生成式人工智能数据风险的管理效果。具体而言，服务提供者需对产品的提供负主要责任；技术研发人员需对数据滥用及算法偏见负责；数据供应者在数据收集时若侵犯他人合法权益，同样需承担侵权责任；系统操作者若因操作不当导致数据泄露等问题，也应承担相应责任。同时，在推动生成式人工智能技术发展的同时，确保数据风险控制得当，准确区分不同责任主体，对于追责生成式人工智能的风险责任至关重要。[16]

4.3 构成生成式人工智能数据风险监管机制。现阶段对于人工智能的监管存在多头管理的问题，治理机构间缺乏协调性。治理分散化不仅阻碍了治理目标的实现，更让治理资源得不到高效分配。监管部门之间往往容易发生冲突，使得监管工作止步不前。在目前多部门合力制定法规的情况下，可以通过权力的合理划配加强各部门之间的监管力度。由网信办牵头，确保其能够在人工智能领域保持统筹地位，对各部门进行统一指导。其他部门在不同治理环节具有不同的责任分工，避免“无人管”、“一起管”职责重合及交叉局面，建立起常态化、规范化的联动机制，以监管适宜促进技术正轨发展，破解多头管理难题。

5. 结语

生成式人工智能技术的蓬勃发展，是我国数字经济的新质生产力，成为助推传统产业转型升级的利器，但是在 ChatGPT、文心一言等 AI 大数据飞速发展的背后，诸多数据风险也随之而来。现行人工智能领域的法律规范在监管主体和内容等方面存在着不足。为进一步推动生成式人工智能的发展，有必要完善人工智能领域的法规，构建协调统一的监管机制，划分不同阶段的责任主体等来缓解生成式人工智能的负面冲击，实现新质技术发展与公众数据安全平衡。

参考文献：

[1] 熊琦、陈子懿：《美国人工智能模型训练合理使用认定的成案经验研究》，载《科技与法律 (中英文)》2024 年第 6 期，第 11–23 页。

[2] 张欣：《论人工智能体的模块化治理》，载《东方法学》2024 年第 3 期，第 129–142 页。

[3] 全燕、张入迁：《新质治理力 : 全球人工智能治理张力下中国倡议的机制创新》，载《传媒观察》2024 年第 11 期，第 14–23 页。

[4] 于水、范德志：《新一代人工智能 (ChatGPT) 的主要特征、社会风险及其治理路径》,载《大连理工大学学报 (社会科学版)》2023 年第 44 卷第 5 期，第 28–34 页。

[5] 张瀚：《人工智能生成内容风险的法律规制研究》，载《网络安全技术与应用》2024 年第 10 期，第 120–123 页。

[6] 陈永伟：《超越 ChatGPT：生成式 AI 的机遇、风险与挑战》，载《山东大学学报（哲学社会科学版）》2023 年第 3 期，第 135–136 页。

[7] 张涛：《生成式人工智能训练数据集的法律风险与包容审慎规制》，载《比较法研究》2024 年第 4 期，第 86–103 页。

[8] 刘永红、李文颖：《生成式人工智能的法律风险及其规制路径》，载《内江师范学院学报》2024 年第 39 卷第 9 期，第 94–100 页。

[9] 武振国：《欧盟人工智能的实验主义治理路径及中国借鉴》，载《西北大学学报 (哲学社会科学版)》2024 年第 54 卷第 6 期，第 153–164 页。

[10] 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》，载《比较法研究》2023

年第 3 期，第 155–172 页。

[11] 朱恒楠：《类 ChatGPT 模型产生的法律风险及应对之策》，载《中国价格监管与反垄断》2024 年第 9 期，第 65–67 页。

[12] 杨金康：《生成式人工智能的法律风险及应对策略》，载《河北企业》2024 年第 8 期，第 155–157 页。

[13] 程韵：《生成式人工智能模型的法律风险及规制研究——以 ChatGPT 为视角展开》，载《网络安全技术与应用》2024 年第 8 期，第 123–125 页。

[14] 杨福忠、姚凤梅：《人工智能赋能社会治理之维度及风险法律防治》，载《河北法学》2022 年第 40 卷第 11 期，第 89–102 页。

[15] 刘琪、于游：《生成式人工智能发展的风险与法律规制》，载《现代商贸工业》2024 年第 45 卷第 15 期，第 223–225 页。

[16] 邓喜莲、刘嘉馨：《人工智能的社会影响和法律规制分析》，载《法制与社会》2021 年第 20 期，第 104–106 页。