

生成式人工智能(GAI)的法律风险初探

作者：杨迅 | 严怡

继风靡全球的 ChatGPT-3.5 之后，美国人工智能研究公司 OpenAI 又推出了 GPT-4。与其前身相比，GPT-4 更新了训练数据，理解能力、措辞水平均有所提高。对于涉及煽动仇恨、骚扰他人等不良内容的请求，GPT-4 基本不予回应。

GPT-4 的安全性、合规性似乎有所提高。然而，GPT-4 问世后不久，意大利数据保护局就以“违法收集数据”为由，暂时将其封禁。继而，加拿大联邦隐私监管局也因 OpenAI 涉嫌“未经同意处理个人信息”而展开调查。

那么，GPT-4 到底为什么屡屡被怀疑侵犯“隐私”？以 ChatGPT 为代表的生成式人工智能(Generative Artificial Intelligence, “GAI”)究竟存在何等法律风险？本文对此作简要探讨。

一. 内容违规风险

作为一种新型生产工具，GAI 的工作逻辑可分为两步：其一，接受巨量数据以训练模型；其二，基于用户的需求，利用模型和其他技术(如：生成语言技术)生成内容。由此可见，通过人工智能生成的内容(即“AIGC”)主要取决于训练数据与用户需求。

基于前述逻辑，进一步地，源于特定文化场域的数据俾使 GAI 偏向于特定价值观；而在此基础上，用户需求将进一步左右 GAI 的判断。以 ChatGPT 为例，普林斯顿大学的研究结果显示：受语言模型中的社会偏见影响，ChatGPT 对特定种族表现出了较大的敌意。而在使用过程中，假使 ChatGPT 被赋予某种负面的人格特质¹，其回答也将含有歧视性偏见、错误的刻板印象等有害内容。²

.....
如您需要了解我们的出版物，
请联系：

Publication@llinkslaw.com

¹ 所谓“赋予某种负面的人物特质”大多指用户要求 ChatGPT 扮演特定人物的情形，如要求其以种族主义者的口吻回答。

² Deshpande, A., Murahari. V., Rajpurohit T., Kalyan. A., Narasimhan K. Toxicity in CHATGPT: Analyzing Persona-assigned Language Models.

同时, GAI 生成之内容, 受到人工输入之素材和指令的影响, 甚至是决定性的当然。于是, GAI 也可能被用于不合法合规的用途。比如, AI 换脸兴起之初, 大量被用于伪造女星不雅图片、视频, 严重侵犯了受害人的肖像权、名誉权及人格尊严。此时, 功能强大的 GAI 不仅没有起到提高生产效率、增进人类福祉的作用, 反而成为了别有用心者开展违法违规行为的便利工具。如果法律不对 GAI 的开发与使用行为加以规制, GAI 终将与我国大力发展人工智能的初心背道而驰。

2023 年 4 月 11 日, 国家互联网信息办公室发布了《生成式人工智能服务管理办法(征求意见稿)》³。该办法对利用 GAI 生成破坏社会秩序与经济秩序的内容的行为予以明确的否定评价, 并要求生成式人工智能的开发者和应用者分别采取措施保障生成内容的合法性, 即:

- (1) 开发人员在算法设计、训练数据选择、模型生成和优化的过程中, 应采取必要措施防止 GAI 生成违法内容;
- (2) 服务提供者一旦发现侵权或者生成违法违规内容的行为, 应立即采取更正、屏蔽、删除等措施, 并进行内容过滤、模型优化训练。

GAI 作为一种工具, 无论如何“智能”, 必然受到开发者、使用者输入的影响, 完全消除违法内容是难以实现的。但是, 从其开发端的算法训练和使用端的素材提供双管齐下, 可以尽可能减小 GAI 生成违法违规内容的风险。

二. 侵犯知识产权

除了类似 ChatGPT 的聊天机器人外, GAI 还衍生出了 AI 绘画、AI 写作等人工智能的新兴应用。譬如, AI 绘画可以通过抓取公开数据, 或利用用户上传的图像, 建立庞大的数据集, 并从中习得“绘画技巧”, 从而生成各种预设风格、题裁的图画。可见, AI 绘画的创作过程, 实际上是汲取在先创作基础上的再创作, 这是 AI 绘画的特色, 也是带来知识产权侵权风险的根源。大批创作者在社交媒体、艺术交流平台发起抵制 AI 绘画的活动, 控诉其侵害著作权。

创作者的担忧确有一定道理。根据《中华人民共和国著作权法》第 10 条的规定, 著作权人享有以印刷、复印、拓印、录音、录像、翻录、翻拍、数字化等方式将作品制作一份或者多份的权利。为收集训练数据, GAI 必定会以数字形式复制作品, 而这显然侵犯了作者的复制权。另外, 部分 GAI 还可能侵犯了作者的改编权。⁴比如, 如果用户选用“图生图”模式, GAI 将基于用户上传的参考图生成图画, 在此过程中, 其有可能侵犯参考图作者的改编权。而在“文生图”模式中, GAI 依据用户的文字性描述生成图画, 并没有直接改变原作品, 可能并不构成侵权。

³ 以下简称“办法”。

⁴ 参见焦和平: 人工智能创作中数据获取与利用的著作权风险及化解路径, 载《当代法学》2022 年第 4 期, 128 页。

更加晦暗不明的是，如果 GAI 的使用者并未上传第三方作品，而是放任 GAI 在互联网上自由地抓取和利用第三方作品，这就导致了 GAI 有可能未经所抓取的作品的作者同意，或者超越作者的授权使用其作品；根据 GAI 不同的处理方式，GAI 创作的作品可能成为原作品的衍生作品，从而构成侵权。

一方面，该种对公开可获得的(并非自由)作品的再利用，根植于 GAI 的基本逻辑，构成侵权风险的根源；另一方面，GAI 所使用的算法是一个“黑匣子”，很难洞察其具体利用现有作品，训练算法，以及生成创作的全过程，给 GAI 使用者侵权行为增加不可控性，也给被抓取的作品的权利人的举证带来困难。

GAI 工具的提供者，也许可以依赖“避风港”原则，减小其风险，即当权利人通知其利用其 GAI 开展侵权行为并提供是当证据的，GAI 提供者应当采取措施(比如停止向被主张侵权人提供 GAI)，避免侵权行为的扩大，在此基础上 GAI 的提供者可以免除民事赔偿责任。但是，对于 GAI 的使用者，其作为直接侵权人(因为，GAI 作为一种工具，不是承担民事责任的主体)，则无法享受“避风港”，无论其是否洞悉其使用的 GAI 的处理过程，都必须对 GAI 生成的作品的侵权承担责任。因而，在使用 GAI 时，了解其内在算法逻辑，评估侵权风险，对于使用者的风险防范，至关重要。

三. 损害数据安全

如前文所言，GAI 须在处理大量数据(可能包含个人信息和非个人数据)的基础上训练成熟的模型。从这一意义上而言，GAI 的开发者和使用者都是数据处理者，应尽到数据安全保障义务。遗憾的是，相对于数据的体量与敏感性，GAI 在数据的安全保障方面表现得差强人意。例如，ChatGPT 的开源库曾出现重大漏洞：用户的姓名、付款地址、信用卡卡号后四位、用户与 ChatGPT 的聊天记录等信息均被泄露，严重损害了用户的个人信息、财产权等合法权益。再如，三星公司的工程师利用 ChatGPT 来修复源代码，反导致新程序的源代码本体、与硬件相关的内部会议记录等机密数据均为 ChatGPT 掌握，成为 ChatGPT 的“数据养料”。可以想见，一旦 ChatGPT 再次接受到类似修复代码的请求，其就极有可能利用三星的机密数据，甚至导致三星部分源代码被利用。

诚然，GAI 是一项基于大数据为用户答疑解惑的技术服务，丰富的数据将使其更加人性化、功能更为强大，但实现这一目的绝不能以破坏数据安全为代价。以三星为戒，企事业组织应当加强对员工的安全教育培训，告诫员工提高保密意识，切勿向 ChatGPT 等 GAI 透露关涉人身安全、财产安全的敏感信息。此外，ChatGPT 的开发人员也应尽快建立完备的数据合规处理机制，其考虑的不应该仅仅是技术可行性，也要将合法合规性纳入设计考量，比如 GAI 接收到个人信息、商业机密及其他重要数据时，除了必要的算法训练外，对原数据立即采取删除、匿名化等措施，降低敏感性。

四. 结论

法律风险的背后往往是纠葛复杂的利益冲突。在科技革命如火如荼的今天，GAI 无疑是一股不容抗拒的潮流，有望成为未来经济发展的重要支撑。但另一方面，GAI 的出现也会对现代经济秩序、社会秩序造成冲击，产生一系列的法律风险。

对于立法者而言，如何平衡技术创新、技术中立和算法正义之间的关系，避免新技术倾轧其他市场主体的利益，这些问题有待社会各界共同探讨、解决。

对于广大 GAI 的开发者和使用者而言，在具体法律制度暂付阙如的状况下，既不能由于法律的不确定性而止步不前，也不能无视法律风险一意孤行，如何根据现有知识产权、个人信息、数据安全等法律原则，动态评估风险，设计合理的商业结构，则是业界和法律专业人士的任务。

如您希望就相关问题进一步交流，请联系：



杨 迅
+86 21 3135 8799
xun.yang@llinkslaw.com

如您希望就其他问题进一步交流或有其他业务咨询需求，请随时与我们联系: master@llinkslaw.com

上海

上海市银城中路 68 号
时代金融中心 19 楼
T: +86 21 3135 8666
F: +86 21 3135 8600

北京

北京市朝阳区光华东里 8 号
中海广场中楼 30 层
T: +86 10 5081 3888
F: +86 10 5081 3866

深圳

深圳市南山区科苑南路 2666 号
中国华润大厦 18 楼
T: +86 755 3391 7666
F: +86 755 3391 7668

香港

香港中环遮打道 18 号
历山大厦 32 楼 3201 室
T: +852 2592 1978
F: +852 2868 0883

伦敦

1/F, 3 More London Riverside
London SE1 2RE
T: +44 (0)20 3283 4337
D: +44 (0)20 3283 4323



www.llinkslaw.com



Wechat: LlinksLaw

本土化资源 国际化视野

免责声明:

本出版物仅供一般性参考，并无意提供任何法律或其他建议。我们明示不对任何依赖本出版物的任何内容而采取或不采取行动所导致的后果承担责任。我们保留所有对本出版物的权利。

© 通力律师事务所 2023