汉坤法律评述
2023 年 2 月 23 日

HANKUN
汉 坤 律 师 事 务 所
Han Kun Law Offices

北京 | 上海 | 深圳 | 海口 | 武汉 | 香港

从 ChatGPT 看生成式 AI 的合规挑战与应对

作者：段志超 | 蔡克蒙 | 蒋海楠 | 邹奕

引言：

2022 年底，由 OpenAI 公司发布的对话式大型语言模型 ChatGPT 一夜爆火，该产品以强大的文字处理和人机交互功能迅速成为热门的新一代人工智能产品，并在不到两个月的时间内用户量突破一亿，其发展速度之快远超以往的任何一个互联网应用。作为生成式 AI，ChatGPT 通过无监督学习学习自然语言的结构和模式，然后通过微调等方式来适应各种具体的任务，同时会使用监督学习和强化学习优化和改进其在特定任务上的表现。与过往其他的颠覆性技术一样，生成式 AI 可能对既有的生产关系、法律制度、社会伦理带来冲击，甚至会影响社会制度的安全和稳定。这些技术在开发、应用的过程中会受到现有法律制度的规范，而法律制度亦在不断革新以应对技术带来的挑战。本文旨在初步探讨像 ChatGPT 这样的生成式 AI 可能产生的合规问题，以及目前法律制度对这些出现端倪的问题做出的应对与规范。

一、生成式 AI 和 ChatGPT

生成式 AI 是指可以生成自然语言文本、图像、音频等形式的 AI 模型。ChatGPT 基于自然语言处理技术，是生成式 AI 在文本领域的应用，其可以通过人类自然对话方式进行交互，还可以应用于相对复杂的语言工作处理，包括自动文本生成、自动问答、自动摘要等在内的多种任务。如：在自动文本生成方面，ChatGPT 可以根据输入的文本自动生成类似的文本（剧本、歌曲、企划等），在自动问答方面，ChatGPT 可以根据输入的问题自动生成答案。ChatGPT 还具有编写和调试计算机程序的能力。ChatGPT 相比以往的主要提升点在于记忆能力，ChatGPT 可以储存对话信息，延续上下文，从而实现连续对话，这在对话场景中至关重要，极大地提升了对话交互模式下的用户体验。与此同时，生成式 AI 在绘画创作、音视频合成、代码编写等领域也同样表现出色。例如，文生图模型 Stable Diffusion，人工智能代码辅助工具 Copilot 在过去的一年同样为用户带来极佳的体验。

生成式 AI 可能带来的颠覆性变革在于过往某些仅能由人类完成的创造性工作可能会被生成性 AI 所取代，或由人机协同创作完成，这会使创造和知识工作的边际成本大大降低，进而大幅提升劳动生产率，创造巨大的经济价值[1]。实际上，生成式 AI 已经在广泛应用于各个领域。根据 Statista 的统计，截止 2023 年 1 月，已经有九百多家企业与 OpenAI 合作，其技术被广泛应用于科技、教育、商业服务、制造、零售、医疗保险、媒体与互联网等多个行业。

[1] 腾讯研究院，《AIGC 发展趋势报告 2023 — 迎接人工智能的下一个时代》。

二、个人信息保护：贯穿生成式 AI 的研发和应用的红线

生成式 AI 需要大量的数据进行训练，并且训练数据的质量和多样性对于模型的性能有重要的影响。此外，其应用场景多元，且类似 ChatGPT 的生成式 AI 更是涉及和终端用户的高频次的直接交互。以上因素都决定了个人信息保护将是生成式 AI 在研发和应用中不可忽视的合规点。

（一）使用公开个人信息

以 ChatGPT 为例，其自称训练数据来源于公开信息，包括网络文本、语言学知识库、对话数据集、科学论文等多种渠道。即使这些渠道的数据可能包含个人信息，开发者也会对数据进行筛选和屏蔽，去除其中的个人信息，仅会使用匿名化数据用于训练。

一些图像生成的生成式 AI 可能在训练过程中使用公开的个人图像信息。需要注意的是，即使是使用公开个人信息，亦可能导致侵犯个人信息权益。例如，美国《伊利诺斯州生物信息隐私法》（Illinois' Biometric Information Privacy Act，“BIPA”）对生物识别信息的收集、使用、交易、披露等做出了严格限制。原则上未经个人书面同意不得处理其生物识别信息。Alphabet、Microsoft、Facebook 等使用公开图片集进行算法训练的均曾深陷 BIPA 诉讼之中[2]。去年 5 月，知名的人脸识别技术公司 Clearview AI 被迫与美国公民自由协会（ACLU）和解，同意限制其人脸数据库访问，停止向美国大多数私营企业和个人提供人脸识别数据库[3]。可预想的，生成式 AI，尤其是文生图式 AI，其为算法训练而进行的对生物识别信息的收集和处理具有相当的合规风险。

（二）个人信息“二次利用”

此外，开发者还可能将在服务中收集的个人信息“二次利用”于算法模型训练的目的。目前 OpenAI 在使用协议[4]中约定用户同意 OpenAI 利用用户的输入及 ChatGPT 的输出结果以改善模型性能，提升用户体验[5]。OpenAI 是否还会将收集的用户数据用于其他场景目前并不清晰。按照《个人信息保护法》规定，基于同意处理个人信息的，个人信息的处理目的、处理方式和处理的个人信息种类发生变更的，应当重新取得个人同意[6]。欧盟 GDPR 亦规定，如个人数据被用于收集之外的其他目的，数据控制者需在“二次利用”前事先告知个人信息主体[7]，且“二次利用”的目的应与原初收集目的兼容。而在企业使用个人信息用于算法训练的情况下，即使算法训练被认为是与原初收集目的兼容，企业通常仍需要重新获得个人同意[8]。即使在一些个人信息保护不那么严格的国家，如新加坡，其个人数据保护委员会（PDPC）在 2020 年修订的《AI 治理框架》（Model AI Governance Framework）中亦建议向用户告知其交互数据将被用于算法研发目的[9]。

---

[2] https://techcrunch.com/2021/03/01/facebook-illinois-class-action-bipa/。

[3] https://www.theverge.com/2022/5/9/23063952/clearview-ai-aclu-settlement-illinois-bipa-injunction-private-companies。

[4] https://openai.com/terms/#:~:text=You%20must%20be%2018%20years,to%20register%20for%20an%20account。

[5] https://openai.com/terms/#:~:text=You%20must%20be%2018%20years,to%20register%20for%20an%20account 另外，如果用户不同意将个人输入输出数据用作模型改进的，Open AI 同样提供不收集用户数据的使用途径。https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance。

[6] 《个人信息保护法》第 14 条。

[7] GDPR, Article 14.3。

[8] 英国信息专员办公室（ICO），《AI 和数据保护的指南（Guidance on AI and data protection）》“*If the data was initially processed for a different purpose, and you later decide to use it for the separate purpose of training an AI system, you need to inform the individuals concerned (as well as ensuring the new purpose is compatible with the previous one)*”。

[9] Model AI Governance Framework 第 3.5 条。

超范围处理个人信息用于算法研发不仅可能带来个人信息合规方面的风险，还可能影响开发者对其算法所享有的权益。例如，美国联邦贸易委员会（Federal Trade Commission，FTC）公布处罚，勒令一家名为“Everalbum”的公司删除其未经用户明确同意收集的照片或视频，以及利用这些数据训练出的所有算法[10]。”

（三）应用过程中的个人信息合规

生成式 AI 在应用过程中同样需要遵守个人信息保护相关法律规定，最为重要的是处理个人信息需具备适当的法律基础。近期，意大利隐私监管机构勒令人工智能聊天应用程序 Replika 停止处理当地用户的数据，原因是 Replika 在向未成年人提供服务同时缺乏处理未成年人个人信息的合法性基础，且与未成年人聊天内容与聊天对象的年龄不相称[11]。此外，生成式 AI 在研发和应用中亦应重视保护个人信息主体的查阅、复制、删除、撤回同意等权利，例如 ChatGPT 即在用户协议中明确，用户有权拒绝将其数据用于模型训练和改进，但这种做法可能影响用户在特定场景下的使用系统能力。

（四）合规建议

我们建议企业在开发生成式 AI 过程中，尽可能使用非个人信息或匿名化数据。如不可避免使用个人信息，应向个人明确告知其数据将被用于算法训练等处理目的并尽可能获得个人同意。如涉及使用公开个人信息，应确保在合理范围内使用公开个人信息，避免对个人造成影响或侵害。此外，企业还应尽可能规避像 BIPA 这样对生物识别信息进行严格保护的法律的适用。

三、训练数据来源的合法性：个人信息维度之外的数据权益考量

生成式 AI 算法训练时所输入的训练数据可能涉及抓取第三方网站的平台数据，或者直接利用他人已加工处理过的数据集合。即使不涉及个人信息，未经授权的收集和使用他人享有权益的数据集合、超过授权范围的使用数据集合可能存在知识产权侵权以及不正当竞争的法律风险，还可能影响开发者对算法或模型的权属。

我国现行法律法规对数据权益的保护正趋于明晰。《民法典》第五章“民事权利”第 127 条上位性的规定“法律对数据、网络虚拟财产的保护有规定的，依照其规定。”《反不正当竞争法（修订草案征求意见稿）》第十八条[12]对商业数据（即经营者依法收集、具有商业价值并采取相应技术管理措施的数据）的获取和使用作进一步规定，明确禁止“以违反诚实信用和商业道德的其他方式不正当获取和使用他人商业数据，严重损害其他经营者和消费者的合法权益，扰乱市场公平竞争秩序”等行为。此外，如果数据集合本身构成《著作权法》下汇编作品的，汇编者对数据集合还享有版权法下的权利，对该等数据的使用需获得汇编者的授权。

司法实践层面，已有法院将数据资源界定为“竞争权益”、“财产性利益”，例如在微信诉群控软件案

[10] https://www.ftc.gov/legal-library/browse/cases-proceedings/192-3172-everalbum-inc-matter。

[11] https://techcrunch.com/2023/02/03/replika-italy-data-processing-ban/。

[12] 《中华人民共和国反不正当竞争法（修订草案征求意见稿）》第十八条：经营者不得实施下列行为，不正当获取或者使用其他经营者的商业数据，损害其他经营者和消费者的合法权益，扰乱市场公平竞争秩序：（一）以盗窃、胁迫、欺诈、电子侵入等方式，破坏技术管理措施，不正当获取其他经营者的商业数据，不合理地增加其他经营者的运营成本、影响其他经营者的正常经营；（二）**违反约定或者合理、正当的数据抓取协议，获取和使用他人商业数据，并足以实质性替代其他经营者提供的相关产品或者服务**；（三）披露、转让或者使用以不正当手段获取的其他经营者的商业数据，并足以实质性替代其他经营者提供的相关产品或者服务；（四）以违反诚实信用和商业道德的其他方式不正当获取和使用他人商业数据，严重损害其他经营者和消费者的合法权益，扰乱市场公平竞争秩序。 本法所称商业数据，是指经营者依法收集、具有商业价值并采取相应技术管理措施的数据。获取、使用或者披露与公众可以无偿利用的信息相同的数据，不属于本条第一款所称不正当获取或者使用其他经营者商业数据。

（〔2020〕浙 01 民终 5889 号）案中，杭州中院认为“就平台数据资源整体而言，系平台投入了大量人力、物力，经过长期经营积累聚集而成的，该数据资源能够给平台带来商业利益与竞争优势，平台对于整体数据资源应当享有竞争权益。”再例如，在淘宝诉美景案（〔2018〕浙 01 民终 7312 号）案中，杭州中院认为“‘生意参谋’数据产品，产品研发者投入大量成本尤其是智力投入，能为其带来可观的商业利益与市场竞争优势，这一数据产品已经成为淘宝公司一项重要财产性权益。”

四、深度合成技术监管：从 DeepFake 到 ChatGPT

AI 生成内容可能造成用户误解或混淆，用于产生虚假信息、垃圾信息，污染网络言论环境，甚至影响政治舆论或政治生态，国内外监管机关在深度合成技术应用问世之初就对其给予了高度关注。为应对人脸替换、语音合成等 Deep Fake（深度伪造技术）技术对法律和秩序的冲击，我国出台《互联网信息服务深度合成管理规定》，旨在规范对深度合成技术的使用。对深度合成技术的监管同样延及类 ChatGPT 的产品和服务。

（一）中国对深度合成技术的监管

我国 2022 年 12 月颁布的《互联网信息服务深度合成管理规定》将“深度合成技术”界定为“利用深度学习、虚拟现实等生成合成类算法制作文本、图像、音频、视频、虚拟场景等网络信息的技术，包括但不限于：（一）篇章生成、文本风格转换、问答对话等生成或者编辑文本内容的技术；（二）文本转语音、语音转换、语音属性编辑等生成或者编辑语音内容的技术；（三）音乐生成、场景声编辑等生成或者编辑非语音内容的技术；（四）人脸生成、人脸替换、人物属性编辑、人脸操控、姿态操控等生成或者编辑图像、视频内容中生物特征的技术；（五）图像生成、图像增强、图像修复等生成或者编辑图像、视频内容中非生物特征的技术；（六）三维重建、数字仿真等生成或者编辑数字人物、虚拟场景的技术。”像 OpenAI 这样的技术开发者可能构成《互联网信息服务深度合成管理规定》项下的“深度合成服务技术支持者”，而依靠 OpenAI 深度合成技术作为驱动的具体服务运营方可能构成《互联网信息服务深度合成管理规定》项下的“深度合成服务提供者”。

《互联网信息服务深度合成管理规定》规定服务提供者的主要义务如下，这些义务的核心在于对生成内容进行有效管控，避免用户混淆或误认，杜绝违法信息和虚假信息传播：

- 对使用其服务生成或者编辑的信息内容，应当采取技术措施添加不影响用户使用的标识，并依照法律、行政法规和国家有关规定保存日志信息；
- 对于“可能导致公众混淆或者误认的”的深度合成服务，“在生成或者编辑的信息内容的合理位置、区域进行显著标识，向公众提示深度合成情况”；
- 加强深度合成内容管理，采取技术或者人工方式对深度合成服务使用者的输入数据和合成结果进行审核；
- 建立健全用于识别违法和不良信息的特征库，完善入库标准、规则和程序，记录并留存相关网络日志；
- 应当建立健全辟谣机制，发现利用深度合成服务制作、复制、发布、传播虚假信息的，应当及时采取辟谣措施，保存有关记录，并向网信部门和有关主管部门报告；
- 发现违法和不良信息的，应当依法采取处置措施，保存有关记录，及时向网信部门和有关主管部门报告；

- 对相关深度合成服务使用者依法依约采取警示、限制功能、暂停服务、关闭账号等处置措施；
- 开发上线具有舆论属性或者社会动员能力的新产品、新应用、新功能的，应当按照国家有关规定开展安全评估；
- 对于涉及提供（i）生成或者编辑人脸、人声等生物识别信息的，以及（ii）生成或者编辑可能涉及国家安全、国家形象、国家利益和社会公共利益的特殊物体、场景等非生物识别信息的相关功能的模型、模板等工具的深度合成服务提供者和技术支持者，《深度合成管理规定》还要求其自行或者委托专业机构开展安全评估。

（二）域外视角

欧盟和美国近年来已提出或通过多部涉及 AI 生成内容治理的法案。欧盟委员会在 2018 年发布了《应对线上虚假信息：欧洲方案》，集中阐释了欧盟委员会面对线上虚假信息挑战的基本观点，提出改进信息来源及其生产、传播、定向投放和获得赞助方式的透明度，改善信息的多样性，提高信息的可信度，制定包容性的解决方案等原则，以实现全面防范视频、图像和文字等虚假信息，避免信息发布者违法操纵舆论等状况[13]。2019 年发布的《可信赖人工智能伦理指引》提出设立人工智能高级专家组，从伦理层面规范深度合成应用。2022 年的《数字服务法》（DSA）更是明确了网络平台（尤其是大型互联网平台）的透明度报告（各类有关透明度的义务，特别是包括提供有关限制信息提供的透明信息（Transparent Information）的义务（例如内容审核义务、推荐系统透明度以及透明度报告的新规则）以及内容审核等义务（例如设置有关举报非法内容和用户或接收方投诉的机制、就针对用户或接收方的不利决定的内部投诉处理程序以及设立非诉讼争议解决机构的正式内容审定要求和程序、规定）。

美国近年来亦提出了一些旨在规范深度伪造的法案，例如 2018 年美国参议院提出的《2018 年恶意伪造禁令法案》（Malicious Deep Fake Prohibition Act of 2018）、2019 年美国众议院提出的《深度伪造责任法案》（Deepfakes Accountability Act）、2019 年 6 月美国两党议员分别在众议院、参议院同时提出《2019 年深度伪造报告法案》（Deepfakes Report Act of 2019）、2020 年《识别生成对抗网络法案》（Identifying Outputs of Generative Adversarial Networks Act）等。这些法案中不乏创新性规定，例如 2019 年《深度伪造责任法案》提出“任何含移动视觉元素的先进技术虚假内容应包含嵌入的数字水印，清楚地标识含更改的音频或视觉内容”，视听内容要求披露不少于 1 份的清晰文字陈述，表明该内容包含被更改的视听内容，并简要说明更改程度；并在该内容的完整时间内，以清晰可读的文本在图像底部进行书面陈述，标识含更改的音频和视觉内容，并简要说明更改程度。音频内容要求在该内容的开头部分以清晰的口头陈述表明包含更改后的音频内容，并简要说明更改程度；如内容超过两分钟，则此后每两分钟内需要有不少于一条附加的清晰口头陈述和附加的简明描述。

五、算法透明度和可解释性要求：将黑箱换成玻璃盒子

生成式 AI 同样需要遵守算法透明度和可解释性要求。中外监管均对此多有关注。在我国，《国家发展改革委等部门关于推动平台经济规范健康持续发展的若干意见》要求“在严格保护算法等商业秘密的前提下，支持第三方机构开展算法评估，引导平台企业提升算法透明度与可解释性，促进算法公平。严肃查处利用算法进行信息内容造假、传播负面有害信息和低俗劣质内容、流量劫持以及虚假注册账号等违法违规行为。”类似的，《互联网信息服务算法推荐管理规定》第 4 条规定，“提供算法推荐服务，应当遵守法律法规，

[13] https://www.secrss.com/articles/17221。

尊重社会公德和伦理，遵守商业道德和职业道德，遵循公正公平、公开透明、科学合理和诚实信用的原则”。第 16 条进一步明确“算法推荐服务提供者应当以显著方式告知用户其提供算法推荐服务的情况，并以适当方式公示算法推荐服务的基本原理、目的意图和主要运行机制等。”

域外监管同样将算法的可解释性作为监管的重中之重。在伦理层面，欧盟发布的《可信人工智能的伦理指南》（Ethics Guidelines for Trustworthy AI）将可解释性作为可信人工智能的四个伦理原则之一，将透明性作为可信人工智能的七个关键要求之一。联合国发布的首个全球性人工智能伦理协议《人工智能伦理问题建议书》（Recommendation on the Ethics of Artificial Intelligence），提出了人工智能系统生命周期的所有行为者都应当遵循的十个原则，其中就包括“透明度和可解释性”。在立法层面，GDPR 第 22 条即旨在重点规制在标准建设。在标准层面，美国国家标准技术研究所（NIST）2023 年 1 月更新的《人工智能风险管理框架》（Artificial Intelligence Risk Management Framework）提出可信算法的基本特征，其中即包括负责任和透明的、可解释和可说明等。然而，算法解释和算法透明度要求在理论和实践层面还面临诸多挑战，例如如何平衡商业秘密保护和算法公开透明、算法问责的要求；算法解释是否必要，是否需要考虑算法应用场景；如何实现可解释性，特别是如何以清晰易懂的语言向普通用户解释复杂问题。

六、歧视、偏见和不良影响：模型和训练数据带来的伦理考量

AI 类技术由模型设立和训练数据带来的歧视和偏见一直被广为讨论，而相关关注在包括 ChatGPT 在内的生成式 AI 上被进一步放大。生成式 AI 的训练数据中可能存在性别、种族、文化和社会偏见，这会导致训练结果输出物中“继承”这种偏见。许多域外立法对此予以了充分关注，例如欧盟正在审议的人工智能法案就强调，应使用高质量的训练、确认和验证数据集，数据集应相关、具有代表性、无误且完整，以最大程度地降低风险和歧视性结果。美国科罗拉多州在立法中提出禁止特定行业（如保险行业）使用 AI 和消费者数据制定歧视性费率（CO S.B. 169）[14]；伊利诺伊等州要求相关企业就纯粹依赖 AI 技术筛选候选人的情况进行数据汇报，以帮助政府部门分析是否存在由于使用人工智能而产生的种族歧视（IL H.B. 53）[15]。

一些算法应用甚至可能对人的主观行为、心理情绪和生命健康造成影响。对此，我国《互联网信息服务算法推荐管理规定》已规定“算法推荐服务提供者应当定期审核、评估、验证算法机制机理、模型、数据和应用结果等，不得设置诱导用户沉迷、过度消费等违反法律法规或者违背伦理道德的算法模型。”欧盟正在审议的人工智能法案更是将（1）在无意识中对人类意识进行操控，从而影响其决定或扭曲其行为，进而对人类造成身体或心理的伤害，（2）利用特定人群因年龄、身体或精神残疾而存在的任何脆弱性，实质性地扭曲与该人群有关的行为，以导致或可能导致其身体或心理受伤害，或（3）公共机构或其代表用以根据自然人的社会行为、已知或预测的个性特征，在一定时期内对自然人的信誉进行评估或分类，导致对某些自然人或整个群体的有害或不利待遇的人工智能系统列为“不可接受风险”的人工智能，禁止开发和应用。在前文分析的 Replik 案件中，意大利监管机关禁止 Replik 的理由之一即人工智能聊天式应用程序 Replik 具有“改善用户的情绪健康，帮助用户理解他们的想法，并通过压力管理、社交和寻找爱情来缓解焦虑”的特性，但该特性也导致针对未成年人和情感脆弱的人的风险增加[16]。

七、知识产权：从输入到产生

[14] https://www.leg.colorado.gov/sites/default/files/2021a_169_signed.pdf。

[15] https://custom.statenet.com/public/resources.cgi?id=ID:bill:IL2021000H53&ciq=ncsl&client_md=cf812e17e7ae023eba694938c9628eea&mode=current_text。

[16] https://ai-regulation.com/italian-dpa-clamps-down-on-replika-an-ai-conversational-agent/。

人工智能机器学习时所输入的训练数据可能属于他人知识产权的客体，数据来源本身可能存在授权不足。此外，人工智能生成物（“**AIGC**”）的可版权性以及权利归属尚不明晰，AIGC 之上是否附着其他财产性权利仍有待讨论。再者，如果 AIGC 独创性不足的，可能还构成对现有作品的侵权，用户和平台方需就侵权责任问题作进一步分配。

- **AIGC 的可版权性以及权利归属尚不明晰。**我国法院对 AIGC 是否属于著作权法下的作品观点不一。在“腾讯诉盈讯”（〔2019〕粤 0305 民初 14010 号）一案中，深圳市南山区法院认为由 Dreamwriter 软件的自动运行并产生的涉案作品可获得著作权法保护，理由是（1）该作品外观上而言符合文字作品的表现要求，具有一定的独创性 （2）作品的生产过程“体现了创作者的个性化选择、判断及技巧等因素”。然而，在在“菲林诉百度”（〔2019〕京 73 民终 2030 号）一案中，北京互联网法院和北京知识产权法院否认为由威科先行库生成的大数据报告构成著作权法意义上的作品，原因是在“自然人创作完成仍应是著作权法上作品的必要条件”，涉案图形的生成过程中没有自然人的参与。实践中，我们注意到 OpenAI 和 Stable Diffusion Public Release 版本都在用户协议中明确声明用户对 AIGC 享用全部权利。

- **AIGC 可能构成对现有知识产权的侵权。**以文生图类 AI 产品为例，AI 基于训练数据以及用户指令所生成的 AIGC 在表达层面与他人作品构成实质性相似时，则存在著作权侵权的问题。如果涉及直接复制他人作品元素的，侵权风险将进一步提高。AIGC 的侵权风险同样表现在商标权、肖像权等领域，例如用户可能指令诱导 AI 生成与注册商标类似的图片或者名人的二创作品。实践中，为降低商业模式整体的法律风险，AI 服务运营方可考虑获得热门 IP 的商业许可，并将商业化活动限定于围绕热门 IP 和已属于公有领域的元素。

- **生成式 AI 平台运营方可能需按规定履行侵权内容治理义务。**生成式 AI 平台运营方属于网络服务提供者，根据《民法典》、《信息网络传播权保护条例》等相关法律法规，其可能负有对侵权行为采取必要措施的义务。但就该“采取必要措施”应作何理解，是否类比于电商平台基于“避风港规则”和“红旗规则”所搭建的全流程的知识产权平台治理模式？我们认为，生成式 AI 本身属于具有信息输入和输出处理能力的“模型”或者“工具”，且将来可能演化成信息网络中的基础设施，传统的“避风港规则”和“红旗规则”所形成的合规义务过于繁重，可能并不适用于生成式 AI 平台运营方。生成式 AI 平台运营方按照云服务提供商标准履行合规义务可能更具有可行性，也更为符合立法本意。

八、其他合规风险

此外，我们注意到生成式 AI 还可能面临以下合规问题：

- **算法备案义务：**《互联网信息服务算法推荐管理规定》要求具有舆论属性或者社会动员能力的算法推荐服务提供者应当在提供服务之日起十个工作日内通过互联网信息服务算法备案系统填报服务提供者的名称、服务形式、应用领域、算法类型、算法自评估报告、拟公示内容等信息，履行备案手续[17]。

- **开源软件合规。**随着开源协作文化的普及和开源软件的大规模使用，生成式 AI 算法中必然集成了

[17] 《互联网信息服务深度合成管理规定》第十九条：具有舆论属性或者社会动员能力的深度合成服务提供者，应当按照《互联网信息服务算法推荐管理规定》履行备案和变更、注销备案手续。深度合成服务技术支持者应当参照前款规定履行备案和变更、注销备案手续。完成备案的深度合成服务提供者和技术支持者应当在其对外提供服务的网站、应用程序等的显著位置标明其备案编号并提供公示信息链接。

众多开源组件。根据 ChatGPT 自述，其在开发过程中就依赖了诸如 PyTorch、Tokennizers 等开源软件和工具来进行开发和训练。企业在利用开源软件时需履行开源许可证中各项义务，否则将被视为未取得权利人有效授权，并进而构成著作权侵权且引发商誉贬损等风险。

- **商业秘密保护。**生成式 AI 在与用户交互中涉及到对商业数据的收集和处理。例如用户可能在类 ChatGPT 产品中输入拟作为商业秘密保护的信息，并进而可能被 AI 模型作为训练素材，在向其他用户提供服务时作为“回答”输出，导致商业秘密泄露。此时 AI 交互服务的提供方可能被追究商业秘密侵权的责任。但同时，我们认为就这种潜在的商业秘密侵权主张，AI 服务提供方存在相当的不侵权抗辩空间，例如，AI 服务提供方可主张用户在向类 ChatGPT 产品中输入相关字段时已知悉相关信息可能会被披露给第三方，因此相关信息不构成商业秘密（需结合服务条款进行抗辩），相关权利基础不存在。

- **出口管制。**根据我国有关技术进出口限制相关法律法规，我国限制部分与人工智能相关技术的出口。2022 年 12 月 30 日，中国商务部官网发布了《中国禁止出口限制出口技术目录（征求公众意见版）》，部分信息处理技术被明确列为限制出口技术，包括“智能汉字语音开发工具技术”、“中文平台技术（中文处理核心技术）”、“中译外翻译技术（机器翻译系统得分>4.5 分）”、“基于数据分析的个性化信息推送服务技术（基于海量数据持续训练优化的用户个性化偏好学习技术、用户个性化偏好实时感知技术、信息内容特征建模技术、用户偏好与信息内容匹配分析技术、用于支撑推荐算法的大规模分布式实时计算技术等）”等。因此，生成式 AI 可能涉及前述技术，若将相关技术向境外提供的，需要向商务主管部门申请《技术出口许可证》，以满足我国出口管制法律法规中的相关要求。

- **答案质量和产品责任。**生成式 AI 并不能保证答案的准确性，如有用户依赖不准确的答案做出判断和决策，可能会导致用户产生损失。对此，ChatGPT 明确在用户协议中对此做出了免责。此外，如果将生成式 AI 作为问答式机器人集成于其他产品和服务的，可能产生答案质量和产品责任的问题。我们注意到 ChatGPT 在答复法律、医学等专业问题时会建议咨询相关专业人士的建议，同时 OpenAI 在用户协议中第的 2.（d）条中已明确规定不对第三方软件通过集成 ChatGPT 服务所提供的产品承担责任。

九、总结

ChatGPT 的火爆加速了国内生成式 AI 的商业化进程，围绕生成式 AI 的生态网络将被可预见地迅速搭建。新的产业形态会激化人工智能领域原本的法律问题，也会引发新的问题与挑战。在这样一个技术与法律交叉的敏感领域，AI 生态参与者在积极拥抱技术的同时，需做好当前法律问题的梳理以及潜在风险点的预案，以合规先行的发展理念确保企业的竞争优势。

特别声明

汉坤律师事务所编写《汉坤法律评述》的目的仅为帮助客户及时了解中国或其他相关司法管辖区法律及实务的最新动态和发展，仅供参考，不应被视为任何意义上的法律意见或法律依据。

如您对本期《汉坤法律评述》内容有任何问题或建议，请与汉坤律师事务所以下人员联系：

段志超

电话： +86 10 8516 4123
Email： kevin.duan@hankunlaw.com

蔡克蒙

电话： +86 10 8516 4289
Email： kemeng.cai@hankunlaw.com