风起云涌的 AIGC: 监管、知识产权与算法安全

作者：杨迅 | 夏雨薇 | 杨蕾

前言

近日，美国科技初创公司 OpenAI 开发的对话式人工智聊天工具 ChatGPT 在全球范围内掀起了一阵“人工智能”热潮。在这一浪潮下，AIGC (Artificial Intelligence Generated Content)技术被推向风口浪尖，一时间，成为时代的宠儿。那么，AIGC 技术的开发和应用会面临什么法律风险，面临哪些法律问题?本文将从以下三部分对 AIGC 的重点问题展开介绍。

上篇: AIGC 之政策与监管

中篇: AIGC 之知识产权

下篇: AIGC 之算法安全

上篇: AIGC 之政策与监管

一. AIGC 及其应用

AIGC(Artificial Intelligence Generated Content)，即人工智能生成内容，是指利用人工智能技术生成的内容，是 AI 在内容创作领域的应用。

一般而言 AIGC 的基本结构可以用四个元素来解释[1]：输入、学习算法、训练算法和输出。输入是将已存在的作品加载到语料库中。输入本质上是训练数据的基本构建块；学习算法利用这些构建块，并通过机器学习算法分析出相关特征，再经过训练算法将输入和学习算法生成的数据链接到输出，最后生成基于算法(包括概率)和用户指示的数据结构形式的有形信息。

AIGC 广泛应用于如在线广告、医学研究和诊断以及面部识别。它基于事实的知识，从这些事实中推导出的规则，进而

如您需要了解我们的出版物，
请联系:

Publication@llinkslaw.com

[1] http://jolt.law.harvard.edu/digest/a-legal-anatomy-of-ai-generated-art-part-i

自主生成创造新的文本、图像、音频、视频、代码、交互内容(如数字人)等各种形式的内容和数据，并包括开启科学新发现并创造新的价值和意义(如 AI 技术应用于新药靶点开发)等。其中一些创意成果已经成功实现商业化，比如由 AI 生成的“the Portrait of Edmund Belamy”在佳士得拍卖会上以 432,500 美元的价格售出。[2]

在我国 AIGC 技术的应用尚属探索阶段。结合近期企业对 AIGC 的关注场景，似乎以下场景将是 AIGC 技术应用的突破口和重点:

第一，金融领域。在金融领域，AIGC 技术可以被应用于智能客服，即通过 AIGC 技术生成数字人，通过数字人集成相关信息资料，为客户提供投资推介、产品查询、互动操作等金融服务；AIGC 技术还可以被用于营销资料的合成，即通过 AIGC 技术自动收集整理金融市场信息，通过既定的算法，对金融市场进行判断，并生成市场营销和推介资料；AIGC 技术还可以构建虚拟金融场景，即通过 AIGC 技术，在虚拟空间构建金融环境和交易场景。

第二，生命健康。在生命健康领域，AIGC 技术可以被用于医学图像处理，即 AI 技术对图像进行分析和处理，实现对人体器官、软组织和病变体的位置检测、分割提取、三维重建和三维显示，对感兴趣区域(ROI)进行定性甚至定量的分析；AIGC 也可以被用于合成虚拟医护陪伴，使用 AIGC 技术生成虚拟医护形象，该虚拟医护还可以通过训练，拥有及时反馈患者疑问，分析患者体征，和应急处理的功能。

第三，娱乐媒体。在娱乐媒体领域，AIGC 可以生成虚拟偶像、游戏中的 NPC 形象，该些形象不再是呆板的，而具有智能互动的功能；AIGC 可以用于影视制作，包括前期的剧本生成，中期的虚拟布景和虚拟动作合成，以及后期的加工、预告片制作、劣迹演员替换等；AIGC 还可以用于协助采访、根据给定的素材生成报道等。

二. 他国对 AIGC 的立法

随着包括 AIGC 在内的人工智能技术的发展和应用，欧美法域也在政策法律层面上对 AIGC 高度关注，近年来出台了一系列有关人工智能的立法。但是，从政策导向上看，美国与欧洲对人工智能的态度具有明显区别。

(一) 美国的人工智能相关政策

美国的人工智能相关政策，旨在继续和维持美国在人工智能领域的领先优势。特朗普政权在 2019 年通过行政命令形式颁布的《美国人工智能倡议(America AI Initiative)》。该倡议要求在以下五个方面支持人工智能发展[3]:

[2] https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html
[3] https://trumpwhitehouse.archives.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/

(1) 研发。该倡议通过指导联邦机构在其研发任务中优先考虑人工智能投资，保持对高回报的基础研发的重点和长期重视。这些投资将加强和提升美国工业界、学术界和政府的研发生态系统，并将联邦人工智能支出优先用于能够直接惠及美国人民的前沿想法。

(2) 数据和计算资源。该倡议指示联邦机构优化的人工智能研究人员和开发人员对高质量数据和计算资源的访问，同时确保隐私和安全。它还支持基于云的 AI 平台和工具的开发。以促进公信力，在保护公民安全、安保、自由、隐私和保密的前提下，提高这些资源对人工智能研发专家的价值。

(3) 标准。联邦机构将通过为不同类型的技术和工业部门的人工智能开发和使用建立标准，促进公众对人工智能系统的信任。该标准将帮助联邦监管机构制定和维护安全和值得信赖的创造和采用新人工智能技术的方法。该倡议还要求国家标准与技术研究所领导制定适当的技术标准，以建立可靠、稳健、可信、安全、可跨平台和可互操作的人工智能系统。

(4) 劳动力：为了帮助教育人工智能研发所需劳动力，该倡议呼吁各机构优先考虑奖学金和培训计划，该倡议指示联邦机构通过学徒、技能计划、奖学金以及计算机科学和加强 STEM 教育、提供培训机会和提高公众对人工智能的认识、获得人工智能相关技能，并支持跨政府、行业和学术界开发 AI 就绪人才管道。

(5) 国际参与：该倡议旨在促进一个支持人工智能研发的国际环境，并为美国人工智能产业打开市场，同时也确保人工智能技术的开发方式符合美国的价值观和利益。

继《美国人工智能倡议》之后，美国还出台过一些有关人工智能的政策，其中比较重要的是 2020 年《国家人工智能倡议法案》(National Artificial Intelligence Initiative Act)，该法案于 2021 年 1 月 1 日作为 2021 财年《国防授权法案》(National Defense Authorization Act)的一部分颁布。《国家人工智能倡议法案》提供了一个跨整个联邦政府的协调计划，以加速 AI 研究和应用并促进国家经济繁荣和国家安全。它还建立了国家人工智能计划，负责监督和实施美国关于 AI 的国家战略。

(二) 欧盟的人工智能相关政策

欧盟秉承其一贯的谨慎态度，在人工智能的发展和应用方面更加关注其可能带来的风险。欧盟希望确保人工智能以人为本、合乎道德、可持续发展，且尊重基本权利和价值观，并且据此尝试将人工智能的应用纳入监管。

欧盟关于人工智能的立法核心是《人工智能法案》(**"AIA"**)，AIA 由欧盟委员会于 2021 年 4 月提出，目前正在由欧洲议会和理事会进行讨论。AIA 试图根据不同类型的人工智能应用程序的风险级别，制定不同级别的规则和义务。

- **《人工智能法案》旨在实现以下四个目标：**

(1) 确保投放到欧盟市场和使用的人工智能系统是安全的，并尊重关于基本权利和欧盟价值观的现有法律。

(2) 确保法律的确定性，以促进人工智能的投资和创新。

(3) 加强对适用于人工智能系统的基本权利和安全要求的现有法律的管理和有效执行。
(4) 促进合法、安全和可信的人工智能应用的单一市场的发展，防止市场分裂。

➢ **《人工智能法案》还要求在欧盟范围内建立统一协调的对人工智能的基本要求：**

(1) 禁止对基本权利或安全造成不可接受的风险(An unacceptable risk)的人工智能系统，比如操纵人类行为、利用漏洞或使用社会评分的系统；
(2) 对基本权利或安全造成高风险(A high risk)的人工智能系统(“**高风险人工智能系统**”)在投放市场或投入使用之前必须满足法律所要求的影响评估要求[4]，包括用于生物特征识别、关键基础设施、教育、就业等的系统[5]；
(3) 对高风险人工智能系统的要求包括：数据质量和可追溯性、技术文档和记录保存、透明度和向用户提供信息、人为监督和干预、准确性、稳健性和安全性[6]；
(4) 高风险人工智能系统必须经过一定的安全评估程序才能投放市场或投入使用，根据系统的类型，该评估可以自行进行，也可以由经官方认可的第三方机构开展[7]；
(5) 高风险人工智能系统的提供方必须先在欧洲数据库中注册他们的系统，然后才能将其投放市场或投入使用[8]；
(6) 高风险人工智能系统的提供方必须向主管当局报告任何严重事件和系统故障[9]，用户必须向提供方或是经销方报告任何严重事件或系统故障[10]；
(7) 特定风险人工智能系统(Certain AI Systems)的透明度义务包括[11]：告知用户他们正在与人工智能系统交互；披露情绪识别或生物识别分类的使用；披露人为生成或操纵的内容；以及
(8) 对基本权利或安全造成极小风险或无风险(a low or minimal Risk)的人工智能系统不受任何特定规则的约束，例如垃圾邮件过滤器等。

三. 我国对 AIGC 的立法

我国对包括 AIGC 在内的人工智能的态度，介于美国和欧洲之间：既从政策层面上鼓励人工智能的发展，又对人工智能，尤其是深度合成技术的开发和使用持相对谨慎的监管态度。

(一) 国家发展人工智能战略

国务院于 2017 年颁布了《新一代人工智能发展规划》(“**AI 规划**”)。作为国家战略，该 AI 规划的目标是到 2030 年使中国成为人工智能研发领域的世界领先者。AI 规划概述了中国人工智能的三阶段发展计划。第一阶段重点发展基础技术和应用，第二阶段重点突破智能感知、推理认知、

[4] Article 6(7) of AIA
[5] Annex III of AIA
[6] Chapter 2 of AIA
[7] Article 43 of AIA
[8] Article 51 of AIA
[9] Article 62 of AIA
[10] Article 29(4) of AIA
[11] Article 52 of AIA

人机协同等关键领域。第三阶段即最后阶段，旨在实现人工智能技术和应用的重大突破，建立产业生态，促进人工智能与 5G、物联网、区块链等其他关键技术的融合。为实现这些目标，AI 规划制定了一系列战略和政策，包括加大人工智能研发投入，促进关键技术和应用发展，加强人才培养和招聘，建立人工智能发展和监管新机制。

国家新一代人工智能治理专业委员会于 2021 年颁布了《新一代人工智能伦理规范》，进一步从伦理和安全角度保障人工智能产业的有序发展。该指南强调了确保人工智能系统安全可靠、保护用户隐私和数据安全以及促进人工智能开发和使用的透明度和问责制的重要性。指南设定了人工智能产业发展的一些基本原则，包括：(1)人工智能系统的设计应确保安全、可靠和可信，并应在部署前经过严格的测试和验证；(2)人工智能的开发和使用应当尊重用户的隐私和数据安全，遵守相关法律法规；(3)人工智能系统应该是透明和负责任的，并就如何做出决策和如何使用数据提供明确的解释；(4)人工智能的开发和使用应遵循公平、正义和人类尊严等伦理原则。

为落实《新一代人工智能发展规划》，系统指导各地方和各主体加快人工智能场景应用，科技部等六部委于 2022 年 7 月颁布了《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》，该意见从打造人工智能重大场景、提升人工智能场景创新能力、加快推动人工智能场景开放、加强人工智能场景创新要素供给等方面出发，旨在探索人工智能发展新模式新路径，以人工智能高水平应用促进经济高质量发展。

同年 8 月，科技部进一步颁布了《关于支持建设新一代人工智能示范应用场景的通知》，该通知明确提出首批支持建设智慧农场、智能港口、智能矿山、智能工厂、智慧家居、智能教育、自动驾驶、智能诊疗、智慧法院、智能供应链 10 个示范应用场景，加快推动人工智能应用，助力稳经济，培育新的经济增长点，进一步强调了深挖人工智能应用场景的重要性。

在地方层面，各主要城市陆续颁布了鼓励人工智能发展的政策。2022 年 8 月，深圳地方人大常委会通过了《深圳经济特区人工智能产业促进条例》，该条例是我国第一部正式的关于人工智能的地方立法。它从政策和资金支持、人才储备、基础设施建设、拓展应用等方面鼓励人工智能产业的发展。上海市人大常委会颁布了《上海市促进人工智能产业发展条例》，尤其突出的是，该条例在鼓励和促进制定人工智能的国家标准、行业标准、地方标准制定中发挥引领作用；并且明确："符合国内领先、国际先进要求的，可以在标准文本上使用人工智能'上海标准'的专门标识。"[12]

(二) 监管现状

我国尚未出台对人工智能的统一立法，但散见的法律法规在各个方面限制和规范了人工智能的开发和应用。该些法律法规包括：(1)2017 年生效的《网络安全法》全面系统地规范网络运行，提出系统安全、信息安全的要求；(2)2021 年生效的《数据安全法》提出了保护数据安全的基本要求，而 AI 技术则是依托于大数据实施的；以及(3)2021 年生效的《个人信息保护法》全面规范个人信息处理行为，尤其是关于利用个人信息的自动化决策，与 AI 技术的应用息息相关。此外，在

12 《上海市促进人工智能产业发展条例》第 37 条

使用个人肖像权深度合成的场景下，我国《民法典》有关个人肖像权的规定给出了一定的指引；我国有关知识产权法律也可用来判断 AI 创造品的知识产权权属。

在对于 AIGC 算法的治理层面，2022 年 3 月国务院网信部门(“**网信办**”)颁布了《互联网信息服务算法推荐管理规定》。该规定：(1)要求算法推荐服务方遵循公开透明原则，公开其算法的基本逻辑、目的和机制；(2)鼓励算法推荐服务提供方综合运用内容去重、打散干预等策略来优化其规则的透明度和可解释性；(3)要求算法推荐服务提供者尊重用户的知情权、选择权和退出权，并为用户提供关闭个性化推荐或删除其个人信息的便捷选项；以及(4)禁止算法推荐服务提供者利用算法从事危害国家安全、扰乱公共秩序、侵犯个人隐私、散布谣言或虚假信息等违法或有害活动。

2023 年 1 月生效的《互联网信息服务深度合成管理规定》则更进一步规范了参与深度合成各方的行为：(1)要求深度合成服务提供者和技术支撑方加强训练数据管理，采取必要措施保障训练数据安全；(2)要求深度合成服务提供者和技术支持者为深度合成产品或服务提供清晰的显著标识；(3)收集或使用其人脸或人声等生物识别信息之前征得用户的同意；以及(4)禁止深度合成服务提供者和技术支持者利用深度合成技术从事危害国家安全、扰乱公共秩序、侵犯个人隐私、散布谣言或虚假信息等违法或有害活动。

我国尤其对于具有舆论属性或者社会动员能力的算法推荐服务提供者，深度合成服务提供者、深度合称服务技术支持者，提出了备案要求。根据《互联网信息服务深度合成管理规定》和《互联网信息服务算法推荐管理规定》的规定，具有**舆论属性或者社会动员能力**(可参考《具有舆论属性或社会动员能力的互联网信息服务安全评估规定》中的适用范围)的算法推荐服务提供者，包括生成合成类、个性化推送类、排序精选类、检索过滤类、调度决策类等算法技术的算法，以及深度合成服务提供者、深度合称服务的技术支持者应当通过互联网信息服务算法备案系统进行备案。

中篇: AIGC 之知识产权

AIGC 创造的作品是否能够获得有关知识产权法律的保护?如果能够得到保护，其权利归属于谁?AIGC 技术的使用和发展给知识产权保护带来怎样的挑战?本节将讨论 AIGC 的有关知识产权问题。

一. AIGC 的知识产权确权问题

由于目前主流观点认为 AI 并不能成为作者，即著作权人。因而，AIGC 能否构成作品取决于其“独创性”以及是否有“人”的参与。

(一) 英国

英国 Copyright, Designs and Patents Act (“**CDPA**”)第 9(3)条规定：***“作者应被认为是为创作作品作出必要安排的人。”***[13]这样看来，CDPA 为 AI 成为作者提供了合法的土壤。但 2020 年 9 月，英国知

[13] Copyright, Designs and Patents Act, 1988, c. 48 $ 9(3) (U.K.).

识产权局发布了一份修改其版权法的呼吁。[14]英国政府认为版权法可以充分保护软件，但如果 AI 生成的作品需要额外的法律保护，那么这种权利的来源就不应该来自于版权法。英国政府提出将版权局限于人类创造的作品的动议的同时，也提出要采取行动(如添加水印)减少人类作品和 AI 作品之间的混淆，避免“虚假归属”(false-attribution)的风险。

(二) 美国

美国版权局则明确拒绝承认 AI 的作者身份并拒绝为 AI 生成的内容提供版权保护。2023 年 2 月 22 日，美国版权局致函《黎明的曙光》(Zarya of The Dawn)的作者克里斯蒂娜·卡什塔诺娃(Kristina Kashtanova)，表示不能对 Midjourney AI 生成的图像进行版权保护，因为虽然绘画是依据作者提供的文本与提示来生成的，但“*用户无法预测 Midjourney 的特定输出这一事实使得 Midjourney 在版权方面与艺术家使用的其他工具不同。*”[15]这一立场与其在 2022 年 2 月以“*人类作者身份是版权保护的先决条件*”为由拒绝注册名为“A Recent Entrance to Paradise”的作品类似。[16]此外，美国版权局实践汇编(Compendium of the U.S. Copyright Office Practices)第 313.2 条也明确指出：“要获得‘作者’的资格，作品必须由人类创造”。[17]

(三) 我国

在我国，使用 AI 创作的作品是否构成著作权法意义上的作品主张存在争议。《中华人民共和国著作权法》第 3 条规定：“*本法所称的作品，是指文学、艺术和科学领域内具有独创性并能以一定形式表现的智力成果。*”因此，在著作权理论下，作品是一种“智力成果”的体现，正如同在猴子自拍版权一案中，版权保护不适用于猴子一样，智力活动是人与动物、机器的根本性区别。

但**在我国司法实践中，已存在认定 AIGC 存在著作权的先例**。在 Dreamwriter 案[18]中，法院肯定了 Dreamwriter 自动生成的财经报道文章系独立创作、在外在表现上与已有作品存在一定程度的差异，具有一定的独创性。其次，Dreamwriter 主创团队在数据输入、触发条件设定、模板和语料风格的取舍上的安排与选择，属于与涉案财经报道文章的特定表现形式之间具有直接联系的智力活动，符合《著作权法实施条例》对“创作”的定义，体现了主创团队的“*个性化的安排与选择*”，而非 Dreamwriter 软件的“自我意识”。因此，Dreamwriter 自动生成的财经报道文章属于我国著作权法所保护的文字作品。

版权保护的存在，是为了鼓励充满活力的创意文化，同时将作品的价值归还给创作者，使他们能够过上有尊严的富足生活，并为公众提供广泛的、负担得起的内容。它旨在保护和奖励创作者和其他权利人在一段时间内对创新所做的努力，故授予其某种“垄断”的权利。它同样也是一

[14] https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property#copyright-and-related-rights
[15] https://copyright.gov/docs/zarya-of-the-dawn.pdf
[16] https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf
[17] https://www.copyright.gov/comp3/chap300/ch300-copyrightable-authorship.pdf
[18] (2019)粤 0305 民初 14010 号

种有效的法律工具，防止肆无忌惮的搭便车行为抑制作品的发展，产生公地悲剧。因此，不宜片面地认为在 AIGC 创作过程中不存在人类的参与而任其进入共有领域，还是要基于“独创性”以及“创作过程”进行个案的分析。

二. AIGC 作品的权属

通过 AIGC 技术生成的作品，如果受有关知识产权法保护，那么权利到底归于 AIGC 的技术开发者，即提供算法的一方，还是 AIGC 的技术使用者，即向 AIGC 提出创作指令的一方?

(一) Nova Productions v Mazooma Games and Others 案

在 Nova Productions v Mazooma Games and Others 案[19]中，探讨了由电脑程序使用位图文件生成的一系列复合帧的版权属于软件开发者还是用户，其中提到“*任何特定屏幕的外观在某种程度上取决于游戏的使用方式。例如，当旋转旋钮转动时，球杆围绕主球旋转。同样地，射击的威力也受到玩家选择按下射击按钮的精准度的影响。玩家不是连续帧图像中创作的任何艺术作品的作者。他的投入不是艺术性质的，他没有贡献任何艺术类型的技能或劳动。他也没有承担创建框架图像任何必要的安排，他所做的一切都是为了比赛。*”

但该案所探讨的作品归属是基于电脑程序而并非 AI 软件生成的。回到 AIGC 的语境中，AI 算法需要大量的语料库来训练算法学习，因此在 AI 软件的开发过程中，还存在其他同样重要的利益相关者，如数据标记者和数据提供者，程序员并非是开发 AI 软件的唯一的主体。

此外，笔者亦尝试与 ChatGPT 互动，在向其提问“是否在回答问题时使用了特定的算法或程序进行分析和决策”时，ChatGPT 回复“*在回答问题时，我并没有使用特定的计算机程序或算法来对您的提问进行分析和决策。我所使用的是人工智能技术，其中包括对自然语言的理解和生成，以及对您提出的问题进行人工分析和推理的能力。……并不是通过预先编写好的程序或算法来完成的。相反，它依赖于机器学习算法和人工智能技术，能够根据您的问题和反馈来不断地进行学习和优化，以提供更加准确和个性化的回答。*”

因此，不同于传统的编程中由软件开发者来编码指令，AI 环境中，软件开发者对于 AI 如何生成作品的控制要少得多。

(二) Dreamwriter 案

在 Dreamwriter 案[20]中，法院认为财经报道文章是由主创团队运用 Dreamwriter 软件完成，并未提及其他参与创作的主体。尤其是，法院认为：使用深度合成技术创作，只不过将风格的确定和素材的选择分离，并不影响作品本身的创造性。因此，文章是原告主持创作的法人作品。但鉴于

[19] Nova Productions Ltd v Mazooma Games Ltd [2006] EWHC 24 (Ch) (20 January 2006).
[20] (2019)粤 0305 民初 14010 号

Dreamwriter 软件的开发者(即设定写作方式和风格选定)和用户(提供素材)皆为一人，在实际上法院并未回答到底由哪一方拥有著作权。进一步地，由于 AI 算法嵌套的非线性结构，AI 模型通常会产生算法黑箱效应，因此，此类算法的可解释性在近年来备受关注。也就是说，并不是 AIGC 的所有逻辑都反映了开发者的技能或判断，因此，在 Dreamwriter 案中的"独创性"论述并不是普遍适用的。

也并不是所有的法院都认为软件开发者有理由成为 AIGC 的所有者。有观点认为，软件用户是 AIGC 的所有者，因为他们为塑造输出的内容提供了大量投入。此外，软件用户的利益更可能受到 AIGC 所有权分配的影响。如在菲林诉百度案[21]中，法院认为"*软件开发者没有根据其需求输入关键词进行检索，该分析报告并未传递软件研发者的思想、感情的独创性表达，故不应认定该分析报告为软件研发者创作完成*"；"*对于软件使用者而言，其通过付费使用进行了投入，基于自身需求设置关键词并生成了分析报告，其具有进一步使用、传播分析报告的动力和预期。*"因此，软件的使用者可以享有相关的权益。

ChatGPT 的公司 Open AI 也认为用户应当享有 AIGC 的所有权益，在用户协议[22]第 3(a)中明确："*在用户遵守用户协议条款的前提下,同意向用户转让其对 ChatGPT 自动生成输出的内容的所有权利、所有权和权益。*"因此，基于意思自治的原理，执行创作行为的用户将会获得 ChatGPT 生成的内容的所有权益。

三. AIGC 的侵权风险

使用 AIGC 技术生成文本、图像或其它内容简单便捷，但也带来了知识产权(尤其是著作权)的侵权风险。

(1) **训练数据:** AIGC 的质量取决于用于训练 AI 模型的数据质量。如果模型在未经许可或适当许可的情况下使用受著作权保护的材料进行训练，则生成的内容可能会被视为侵权。

(2) **缺乏原创性:** AIGC 本质上是通过算法学习获取素材的信息，再根据用户操作生成新的内容。理论上，它合成的内容可能与现有版权材料非常相似或存在相同的内容。如果生成的内容不是原创的，则可能被视为现有的第三方作品的衍生作品并可能侵犯原创作品的版权。

(3) **合理使用:** 合理使用是一种法律原则，允许在某些情况下即使未经许可，公众仍可在合理范围内使用受版权保护的材料，例如基于批评、评论、新闻报道、教学、学术或研究的目的。但是，特定的使用是否合理，认定过程可能很复杂且因具体情况而异。在中国，AIGC 源于他人作品时，可能不属于合理使用的范围，即使该种使用基于教育或非商业目的。

(4) **署名:** 《著作权法》赋予了作者署名的权利。AIGC 通常并不能准确指明参考或利用的第三方文献的作者，导致侵犯原作者的署名权。

[21] (2018)京 0491 民初 239 号

[22] 3.(a) "subject to your compliance with these Terms, OpenAI hereby assigns to you all its right, title and interest in and to Output. OpenAI may use Content as necessary to provide and maintain the Services, comply with applicable law, and enforce our policies."

四. AIGC 有关的商业秘密失密风险

使用 AIGC 技术，可能会增加商业秘密泄露的风险，因为 AIGC 技术可能会根据可能包含敏感信息(包括商业机密)的大量数据进行训练。如果 AIGC 系统设计或实施不当，可能会在不经意间泄露或曝光商业秘密。商业秘密泄露的主要方式之一是通过所谓的“模型反转攻击”。该攻击方法主要利用机器学习模型的置信度来反推训练数据，以从底层训练数据中识别和提取敏感信息或商业机密。

另一个风险是 AIGC 系统可能会生成无意中泄露机密信息或商业秘密的文本或内容。例如，基于大型客户评论数据库训练的 AIGC 系统可能会生成无意中泄露有关公司产品或服务的敏感信息的文本。

据报道[23]，亚马逊的公司律师警告员工不要与 ChatGPT 分享“*任何亚马逊的机密信息(包括你正在编写的亚马逊代码)*”。“*这很重要，因为你的输入信息可能会被用作 ChatGPT 进一步迭代的训练数据，我们不希望它的输出包含或类似于我们的机密信息。*”

因此，企业需要妥善建立商业秘密保护制度，通过管理和技术手段，控制 AIGC 技术的滥用导致商业秘密的泄露。

下篇: AIGC 之算法安全

ChatGPT 与其他类似的智能工具，通过深度学习技术等机器学习算法，处理和分析大量的自然语言任务，对自然语言进行解构与生成，并不断地像“人类”一般，在对话中调试、优化原本的算法，以便能够更好满足用户对话需求。

可见，AIGC 技术的核心在于算法。AIGC 的内容是自动生成的，但是算法是人为的。算法的安全性、可靠性决定了 AIGC 的安全性和可靠性。

一. 算法的可靠性考量

AIGC 基于来自公开领域的训练数据，基于模型对大量数据的学习和推理，并结合模型对用户需求的理解，不但能保证生成速度与内容丰富性，还能确保其输出观点满足社会性的要求，能在极短时间内产出大量研究型、检索类的工作成果。

技术先进并不意味着其不存在局限。

首先，AIGC 存在“Garbage in Garbage out”的困境。公开领域的数据良莠不齐，加入模型训练的数据中，无法避免存在事实性的错误、谣言、虚假图片等。在训练数据存在问题的情况下，难以确保模型训练结果准确。

[23] https://baijiahao.baidu.com/s?id=1756596310735491890&wfr=spider&for=pc

其次，即使假定公开领域的数据准确，其也未必客观，而训练后模型生成的结果也可能存在偏向。以互联网为例，社会学理论“沉默的螺旋”可以解释为何不同平台的言论倾向截然不同的现状——即，在一个群体内，观点互相认可的人会不断发言促成讨论，而观点不同的小众人群，会倾向于沉默，最终在群体中只有一种主流声音。互联网的发展本身也是更高维度的“沉默的螺旋”，网络主流受众的声音越大，而借助传统媒体、渠道发声的人声音越小。自此，即使训练数据来自公开渠道，也不可避免可能在某一因素上存在社会性的偏差与误解。

此外，模型是人类工程师训练的产物，在人类工程师介入搭建、调试算法模型的过程中，还可能因为人类的因素导致偏向，例如性别、阶级、乃至社会意识形态上的。一个典型案例是，研究员对 Dalle 2(一个智能图像生成 AI)模型进行基于偏见陈述等领域的定性调查时，发现模型存在“强化刻板印象”的倾向，且这种偏见来源于系统设计的多个环节，甚至与监控团队工作人员的背景和履历的局限性相关[24]。最终导致的结果是，Dalle 2 预览生成图像更倾向于西方白人面孔，输入“私人助理”、“护士”，极大可能生成女性图像，特别是少数族裔女性。提到“律师”“CEO”则生成大量白人和男性穿着西式服装的图片。

可见，虽然 AIGC 技术是技术中立的，但是由于算法和素材的人为干预，其生成的内容并非一定是中立的。在技术中立性原则的前提下，不可避免地需要对技术使用效果进行评价与把控，并从结果监测上反推并调整模型的缺陷。

中国的司法实践中，也对使用算法的局限性有着深刻的认识，并认为企业在使用相应先进算法机制的时候，有更高的注意义务防止相应算法产生不良后果。

在 2022 年宣判的“算法推荐第一案”中北京爱奇艺科技有限公司诉北京字节跳动科技有限公司(以下简称“字节公司”)侵害《延禧攻略》信息网络传播权一案，法院对字节公司使用信息流推荐算法传播侵权作品的行为进行了如下评价：“*字节公司以其服务特点和技术优势…，…也为自身获取了更多的流量和市场竞争优势等利益。但……也存在着提高侵权传播效率、扩大侵权传播范围、加重侵权传播后果的风险。……与不采用****算法推荐、仅提供信息存储空间服务的其他经营者相比****，理应对用户的侵权行为负有更高的注意义务。*”[25]此外，法院也进一步强调了，虽算法机制本身无法直接用于筛选侵权短视频，但算法并不只是互联网企业服务中的全部，相应的算法机制的使用者仍可以“*其****服务和运营的相应环节****中施以必要的注意、采取必要的措施加以完善。*”[26]

二. 算法透明原则要求的实践与意义

通常而言，AIGC 技术的算法是技术开发者的核心资产和重要的商业秘密，是不为公众所知的黑匣子。但是，另一方面，当 AIGC 产生对公众有影响的内容时，其算法本身就不得不落入公众视野，受到公众质疑，以确保其可靠性。

[24] 《DALL·E 2 Preview - Risks and Limitations》 https://github.com/openai/dalle-2-preview/blob/main/system-card.md?utm_source=Sailthru&utm_medium=email&utm_campaign=Future%20%20%20Perfect%204-12-22&utm_term=Future%20Perfect#explicit-content
[25] (2018)京 0108 民初 49421 号案
[26] (2018)京 0108 民初 49421 号案

(一) 算法透明原则

除了对技术使用效果进行管理以外，各国也逐渐对技术本身进行监管。其中保持决策合理公平、实现算法问责的主要机制之一即为“算法透明原则”——要求算法所有者在一定程度上对算法的机制、决策过程或对个人权益影响的情况进行披露、公示，旨在数据隐私、消费者权益保等领域保护公众知情权。

实现算法透明的技术难度在于，如《中篇：AIGC 之知识产权》所述，机器学习算法存在“算法黑箱效应”。当前机器学习技术称为“深度学习”，其基于通过非线性的神经网络提供海量的数据，再经过巨大的神经网络处理该些数据，导致了算法的推演并不完全依人类逻辑，故有些部分无法被人们翻译解读，也就造成了其部分推演无法被完整解释的情况。[27]从法律层面讨论，不断调试且进化的算法模型可能构成商业秘密或其他能够受到保护的客体，对其进行过分披露会侵害公司的权益，损害公司的竞争优势。

(二) 境外算法透明原则的实践

目前，各法域面对“算法透明原则”也在逐步发展。General Data Protection Regulation (“**GDPR**”)下，当数据控制者进行了完全自动化的决策程序，且给数据主体带来了法律后果或其他类似的显著后果时，应当披露以下内容：

- 自动化的决策程序(包括用户画像)的存在；
- 其中所涉逻辑的有意义信息；以及
- 对数据主体的意义以及预期后果[28]

在解释何为“其中所涉逻辑的有意义信息”时，欧盟 WP29 工作组在针对个人自动化决策和用户画像的相关指引中解释，该等披露并不要求是对算法的披露，或是对算法的复杂描述[29]。GDPR 的 Recital 58 也强调，实践中的技术复杂性以及角色的不断增加，数据主体难以理解数据处理的情况，此时透明性原则尤为相关[30]。故，GDPR 所要求的“**有意义信息**”是以数据主体作为导向的，可以认为，有意义信息的衡量标准是，是否能够让数据主体足够了解决策的原因[31]。

---

[27] Big data, artificial intelligence, machine learning and data protection, ICO, Sept, 2017 https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

[28] **General Data Protection Regulation Article 13(2)(f), 14(2)(g), 15(1)(h)** …provide the data subject with the following further information necessary to ensure fair and transparent processing: the existence of automated decision-making, including profiling, ……, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

[29] Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679

[30] **GDPR Recital 58** This is of particular relevance in situations where the proliferation of actors and the technological complexity of practice make it difficult for the data subject to know and understand whether, by whom and for what purpose personal data relating to him or her are being collected, such as in the case of online advertising.

[31] **The WP29 Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679** The information provided should, however, , be sufficiently comprehensive for the data subject to understand the reasons for the decision.

而在解释与说明算法的意义以及预期后果时，欧盟 WP29 工作组认为控制者可以通过使用工具(例如图表)、举例假设等的方式让自己的说明内容更为具象化。GDPR 也考虑到了算法透明的局限性，在 Recital 63[32]中明确，基于透明原则做出的披露不应对商业秘密以及其他知识产权有负面影响。

美国加州于 2023 年正式生效的隐私权法案(The California Privacy Rights Act, "**CPRA**")，目前暂未对与隐私有关的算法进行直接规制，但是 CPRA 条文中已经保证将把"*颁布规范以规制自动化决策工具，包括用户画像以及**要求企业向个人提供与其中所涉逻辑的有意义信息***"纳入 CPRA 整体的法规制定计划中[33]。此外，美国也在 2022 年公开了更新的《算法问责法案》(Algorithmic Accountability Act of 2022)的草案，提出了企业在进行自动化决策、以及重大的关键决策程序[34]时，应当进行相应的影响评估，并向相关的机关提供该等影响评估文件的要求。影响评估中应当分析对消费者可能造成的重大负面影响与改善措施，以还覆盖了对系统的当前与历史性能的测试和评估，以及隐私保护、数据安全措施、决策的公平性、非歧视性等方向的评估。

(三) 我国算法透明实践

如《上篇：政策与监管》所述，在我国具有***舆论属性或者社会动员能力***的算法推荐服务提供者，需要通过互联网算法备案系统进行备案。在备案中，申请备案方需要提交"拟公示内容"，对自己使用的算法的基本原理和运行机制进行解释。

我国《互联网信息服务算法推荐管理规定》同时要求，所有算法推荐服务者应当以***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***。

除了根据法律法规的要求外，在个人信息保护领域，通过计算机程序对用户进行分析与决策的，本身也需要做出类似披露。《个人信息保护法》下，自动化决策指"指通过计算机程序自动分析、评估个人的行为习惯、兴趣爱好或者经济、健康、信用状况等，并进行决策的活动。"(《个人信息保护法》第 73 条)；个人信息处理者应当以显著方式、清晰易懂的语言真实、准确、完整地向个人告知……***个人信息的处理目的、处理方式，处理的个人信息种类、保存期限***（《个人信息保护法》第 17 条）；个人信息处理者进行自动化决策的时候，应当保证决策的透明度、结果公平、公正。对个人权益有重大影响的决定，***个人有权要求个人信息处理者予以说明***，并有权拒绝个人信息处理者仅通过自动化决策的方式作出决定。(《个人信息保护法》第 24 条)

[32] **GDPR Recital 63** That right should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software.

[33] **CPRA 1798.185 (a)(16)** On or before July 1, 2020, the Attorney General shall solicit broad public participation and adopt regulations to further the purposes of this title, including, but not limited to, the following areas: …Issuing regulations governing access and opt-out rights with respect to businesses 'use of automated decisionmaking technology, including profiling and requiring businesses' response to access requests to include meaningful information about the logic involved in those decisionmaking processes, as well as a description of the likely outcome of the process with respect to the consumer.

[34] **Algorithmic Accountability Act of 2022 Section 2(1)** The term "augmented critical decision process" means a process, procedure, or other activity that employs an automated decision system to make a critical decision.

此外，自动化决策的行为本身就是法律规定的事前进行个人信息保护影响评估的情形之一(《个人信息保护法》第 55 条)事前评估的其中一个参考项即是***是否向用户说明了自动化决策的基本原理或运行机制***(《信息安全技术 个人信息安全影响评估指南》附录 A.7)。

目前，据公开资料显示，已经有两家生成合成类算法技术的 AI 智能客服(“菜鸟物流智能客服算法”、“天猫小蜜智能客服算法”)在互联网信息服务算法备案系统中进行了备案。从目前公示的内容来看，备案方对算法运行原理和机制的解释较为简单，主要强调其未对用户的个人信息进行使用或脱敏后方进行使用，且倾向于解释智能客服的用途和功能。

此外，对于较为热门的“个性化推荐”类算法，即通过处理个人信息从而生成个性化信息流、推送等的算法机制，每家企业的公示的侧重点皆有不同。“淘宝推荐算法”“抖音个性化推荐算法”“大众点评首页推荐算法”旨在强调，其提供了相应的技术方案以防个性化推荐类算法下“信息茧房”“信息同质化”负面效果的发生。相反地，“网易传媒信息推送算法”“网易传媒信息流推荐算法”则对其使用的“内容标签系统”“用户画像系统”，推荐机制中的“召回”“过滤”“粗排”“精排”“重排”流程进行了功能性的详细描述。而“微博个性化推送算法”也详细解释了若干算法机制，并辅以**流程图**的方式帮助用户理解微博的个性化推送算法机制。此外，还有部分公示内容中，例如“淘宝推荐算法”还提供了用户关闭个性化信息展示的路径。也可以给落入备案要求及公示要求中的企业一些实践的参考。(以上请见 https://beian.cac.gov.cn/ 公示内容)

结语

2023 年 1 月，微软宣布正和 ChatGPT 开发团队 OpenAI 进行洽谈，计划将其整合到云服务、搜索引擎、甚至 office 中。AIGC 软件正从娱乐软件迈向效率工具，带来的知识产权问题不容忽视。互联网算法备案系统正式实施，意味着我国是国际上较早对算法机制进行明确规范和监管的国家之一。而无论是备案、还是公示要求，用户都需要明白，算法透明本身并非目的，而不过是基于对技术使用效果的审慎态度、对个人权益的平衡、对公共安全与国家安全的考量后的调整工具与手段。

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	北京	深圳
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	伦敦	
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	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。

© 本篇文章独家授权威科先行法律信息库发布，未经许可，不得转载。