LAW, STRATEGY, AND GAME THEORY IN THE AGE OF INTELLIGENT REVOLUTION

Artificial Intelligence

volume 3.0

人工智能3.0：智能浪潮下的法律、博弈与战略

中伦研究院出品

- AI立法与监管：全球纵览
- AI之知识产权布局
- 产业治理热点：AI潜在风险
- AI与地缘困局：贸易合规突围
- AI+律师新范式：行业变革

中伦 ZHONG LUN

中伦研究院出品

contents 目录

contents 目录

preface 前言

人工智能（Artificial Intelligence, AI）作为引领全球新一轮科技革命与产业变革的战略性技术，正以前所未有的速度、广度和深度重塑生产函数与社会关系的基础架构，对全球经济发展模式、社会治理结构及人类文明进步产生深远影响。语言大模型、多模态大模型、智能体以及具身智能等方向持续取得突破性进展，推动人工智能技术从专用领域向通用智能阶段稳步演进。与此同时，技术的工程化与落地应用不断加速，新产品、新服务与创新模式接连涌现，与金融、制造、医疗、教育等各行业的结合日渐走深向实。人工智能正展现出强劲的动力，持续为经济与社会的创新发展注入关键动能。

近年来，无论是在全球范围，还是在中国，AI技术的商业化速度和普及均创历史新高。从全球维度来看，2025年全球人工智能市场规模估值为3909.1亿美元，预计到2033年将达到 34972.6亿美元，2026 年至2033年的复合年增长率（CAGR）高达30.6%[1]；从中国市场来看，作为全球AI发展的重要阵地，2024年中国人工智能产业规模已超9000亿人民币，同比增长24%，[2]2025年中国人工智能**核心产业**规模预计将历史性地**突破1.2万亿**人民币[3]。在企业应用层面，AI技术的渗透率持续攀升，2025年全球约88%的企业在至少一个业务职能中常态化使用AI技术，相比2024年的78%**显著提升**[4]。这种技术渗透速度不仅推动了经济与生产力变革，也带来了前所未有的风险和法律挑战，2019至2024年间，全球记录在案的人工智能风险事件由约400件跃升至7900余件，总量增长近 20 倍[5]，涵盖技术滥用、算法偏见、数据安全、跨境合规等多重维度。

1.Grand View Research, "Artificial Intelligence Market Size, Share & Trends Analysis Report (2026 - 2033)", 2025, https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market.
2.《中国信通院：2024年我国人工智能产业规模超9000亿元》，人民网，2025年9月24日，http://finance.people.com.cn/n1/2025/0924/c1004-40571272.html.
3.中华人民共和国国务院新闻办公室（China SCIO）：《Full Power to AI with Stronger Infrastructure》，2025年12月15日，http://english.scio.gov.cn/chinavoices/2025-12/15/content_118227982.html.
4.McKinsey & Company and QuantumBlack,"The state of AI in 2025: Agents, innovation, and transformation", https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai.
5.世界互联网大会人工智能专业委员会安全与治理推进计划：《为人类共同福祉构建全球人工智能安全与治理体系》报告，2025年11月8日于世界互联网大会乌镇峰会发布。报告引用的数据来自经济合作与发展组织"人工智能风险事件和危害监测器（OECD AIM）"。

中伦长期深耕人工智能法律领域，持续追踪行业动态与规制演进，研究院已于2021年、2023年分别推出《中伦人工智能知识产权保护和数据合规》《人工智能2.0：全景透视AIGC的法律挑战与合规路径》，本文集为系列第三部。文集立足全球视野与实务需求，以 “解析规则、防控风险、赋能实践”为核心宗旨，介绍了全球人工智能立法概览，聚焦中、欧、美等主要法域的最新监管规则，梳理核心地域的合规要求与监管趋势；针对地缘政治背景下AI领域的跨境贸易、算力获取、投资限制等热点问题，提供应对思路；围绕AI 模型、训练数据、生成内容等核心资产的法律保护问题，给出明确指引；关注AI智能体、深度伪造等前沿应用带来的法律挑战，提供前瞻性解决方案；聚焦法律行业自身的AI应用，从保密义务角度出发，探讨合规部署路径；此外，书中还附有实用的“人工智能企业合规义务清单”，可供企业直接对照使用。

本文集旨在为相关主体搭建认知 AI 的监管背景与法律规则、交流实务经验的专业渠道，助力企业在合规框架内高效把握AI技术赋能的发展契机，推动人工智能产业在法治轨道上实现规范化、高质量发展。未来，我们将持续追踪人工智能行业的前沿动态与规制演进，不断输出更具专业深度与实践价值的法律合规成果，为人工智能产业的健康有序发展提供坚实法治保障。

chapter 01

人工智能之立法与监管｜全球概览

artificial intelligence legislation and regulation : a global overview

作 者 / 蔡鹏

发展与安全的双轮驱动：中国人工智能立法演进与治理前瞻

人工智能（AI）作为引领新一轮科技革命和产业变革的战略性技术，已成为重塑全球经济结构、改变国际竞争格局的核心力量。在这一时代背景下，世界主要国家纷纷将发展人工智能提升至国家战略高度。中国明确提出，到2030年要成为世界主要人工智能创新中心，这一宏伟目标为中国人工智能领域的立法与治理实践提供了根本动力。从中国的人工智能治理路径上分析，可以发现其并非一次性的、静态的立法行动，而是一个动态演进、持续迭代的战略过程。

本文旨在梳理中国人工智能立法的演进脉络、核心特征与未来趋势。笔者理解，中国的人工智能治理体系体现了一种独特的“双轨并行”战略——在国家层面强力推动技术创新与产业发展的同时，同步构建日益全面的风险管控与安全保障机制。这一路径从宏观的顶层战略规划起步，逐步演进至针对高风险应用场景的“小切口”式精准规制和敏捷监管，目前正朝着构建一部统一、综合、以平衡风险为基础的根本性法律框架迈进。

为全面展现这一复杂图景，本文将分为三个章节。第一章将追溯中国人工智能政策与立法的历史演进，勾勒出其从战略设想到具体规制的清晰脉络。第二章将聚焦司法前沿，通过分析一系列典型案例，揭示司法机关在填补立法空白、探索规则边界方面所扮演的关键角色。第三章将立足未来立法需求，尝试围绕六大重点议题开展深度前瞻分析。

001> 中国人工智能立法的演进脉络与核心特征

中国的人工智能立法进程并非一蹴而就，而是遵循着一条从宏观到微观、从原则到规则、从鼓励发展到规范发展的清晰轨迹。这一演进过程大致可分为三个相互关联、层层递进的阶段：顶层设计与战略起步、聚焦具体风险与场景规制，以及迈向综合治理体系。

1.1 顶层设计与战略起步 (2017-2021)：宏大战略奠基

中国人工智能治理的起点，可以追溯至2017年7月甚至更早。彼时，国务院发布的《新一代人工智能发展规划》，该规划系统性地提出了面向2030年的"三步走"战略目标，旨在抢占人工智能发展的全球制高点。它不仅明确了技术研发、产业升级和人才培养等方面的具体任务，更重要的是，它为后续所有相关政策和立法活动奠定了基调——即以国家力量驱动人工智能的跨越式发展。这一阶段的政策文件的核心特征在于其宏观性、前瞻性和激励性，重点在于调动全国资源、确立技术雄心，而非施加具有约束力的法律义务。

与产业雄心并行的是对伦理边界的初步探索。在人工智能技术可能带来的社会伦理挑战日益显现的背景下，中国开始着手构建伦理规范的"软法"基础。2021年，全国信息安全标准化技术委员会（TC260）发布的《网络安全标准实践指南——人工智能伦理安全风险防范指引》以及国家新一代人工智能治理专业委员会发布的《新一代人工智能伦理规范》，是这一时期的代表性成果。这些文件首次系统性地提出了"以人为本""智能向善""安全可控""公平公正"等核心伦理原则，并强调将伦理考量贯穿于人工智能研发、设计、应用的全生命周期。它们虽然不具备强制法律效力，但为整个行业树立了价值导向，也为后续的硬法制定提供了理论基础和原则共识。

总体而言，这一初始阶段体现了典型的国家主导型产业政策思维。其首要目标是为新兴产业的蓬勃发展创造最有利的环境，通过战略引领和伦理倡导，在"画跑道"和"立灯塔"的层面进行布局，为后续更精细化的法律规制预留了充分的空间。

1.2 聚焦具体风险与场景规制（2022-2025）："小切口"敏捷治理模式

随着人工智能技术，特别是生成式人工智能的爆发式发展，其潜在风险也迅速浮出水面。虚假信息、算法歧视、个人信息滥用、知识产权侵权等问题，对公共利益和个人权益构成了直接威胁。面对这些迫在眉睫的挑战，中国的治理策略从宏观规划转向了更为精准、务实的"小切口"式立法。

这一阶段的标志性特征是，监管机构针对特定技术、特定场景，快速出台了一系列部门规章和规范性文件。2022年实施的《互联网信息服务算法推荐管理规定》旨在解决"大数据杀熟"和信息茧房等问题。2023年实施的《互联网信息服务深度合成管理规定》则直接剑指"深度伪造"（Deepfake）技术滥用带来的风险。

其中，最具里程碑意义的是2023年7月由国家网信办联合七部门发布的《生成式人工智能服务管理暂行办法》。作为全球首部专门针对生成式AI的国家级法规，它集中体现了中国"发展和安全并重"的核心治理理念。该《办法》一方面鼓励技术创新，另一方面划定了明确的法律红线，如要求"提供具有舆论属性或者社会动员能力的生成式人工智能服务的"主体进行安全评估和算法备案，服务提供者对训练数据来源的合法性负责、对生成内容进行显著标识、建立健全用户投诉机制等。值得注意的是，其"暂行"的性质，恰恰反映了一种敏捷治理（Agile Governance）的智慧：在技术快速迭代的背景下，先通过一部临时性法规迅速应对最突出的风险，在实践中积累监管经验，为未来制定更成熟、更稳定的法律奠定基础。

与这些法规相配套的，是一系列国家标准的密集出台。全国网安标委发布的《生成式人工智能服务安全基本要求》（TC260-003-2024）、《网络安全技术 生成式人工智能预训练和优化训练数据安全规范》（GB/T 45652-2025）等技术文件和标准，将法律法规中的原则性要求，转化为可度量、可验证的技术指标。这种"法律+标准"的双轮驱动模式，构成了中国AI治理的一大特色，确保了监管要求的可落地性。

1.3 迈向综合治理体系：从技术框架到统一法典

在通过"小切口"立法积累了丰富的实践经验后，中国的人工智能治理开始迈向一个更为系统化、体系化的新阶段。其目标是整合前期分散的规则，构建一个统一、协调的综合性法律框架。

这一阶段的预热之作，是全国网安标委于2024年9月发布、并于2025年9月迅速迭代至2.0版的《人工智能安全治理框架》（简称"**框架2.0**"）。它首次系统性地提出了一个基于风险管理的治理方法论，将人工智能安全风险划分为**技术内生安全风险**（如算法偏见、模型缺陷）、**技术应用安全风险**（如网络攻击、内容安全）和**应用衍生安全风险**（如伦理冲击、社会影响）三大类别。在此基础上，框架2.0倡导实施**风险分级管理和敏捷治理**，即根据应用场景、智能化水平和影响范围等维度对风险进行科学评估，并采取与之相适应的、差异化的治理措施。这套完整的理论体系，有利于为未来统一的《人工智能法》提供有效的立法逻辑支撑。

备受瞩目的《人工智能法》草案的立法进程，也反映了这种从审慎到成熟的演变。该草案曾被列入《国务院2024年度立法工作计划》的预备审议项目，但在《全国人大常委会2025年度立法工作计划》中，其表述调整为"由有关方面抓紧开展调研和起草工作，视情安排审议"。这种调整并非立法的停滞，而恰恰是一种战略性的审慎。它表明，立法机关正在充分吸纳前期各项暂行规定、治理框架和司法实践的经验，力求最终出台的法律能够经得起技术发展和实践应用的考验。

笔者认为，这种独特的演进路径揭示了中国AI治理的一个深层逻辑：一个"规制-学习-整合"的迭代循环。首先，面对新兴风险，通过"小切口"的暂行规定进行快速反应和压力测试，这相当于在真实世界中建立监管的"试验田"。其次，从这些"试验田"的实践中学习，识别出真正的监管难点、法律漏洞和行业痛点。最后，将这些经过实践检验的经验和教训，系统性地整合、提炼并升华为理论体系，并最终用法典化的形式固化在统一的《人工智能法》之中。因此，未来的《人工智能法》将不会是空中楼阁，而是深深植根于中国本土实践的、一部高度务实和具有前瞻性的法律。

为清晰展示这一演进过程，下表梳理了中国人工智能领域的关键立法与政策里程碑。

表1：中国人工智能关键立法与政策里程碑

生效日期	文件名称	发布机构	核心贡献与意义
2017年7月	《新一代人工智能发展规划》	国务院	确立国家AI发展的"三步走"战略，是所有后续政策的顶层设计，标志着国家战略的全面启动。
2021年9月	《新一代人工智能伦理规范》	国家新一代人工智能治理专业委员会	提出了将伦理融入AI全生命周期的系统性指引，为行业发展确立了"智能向善"的价值基石。
2022年3月	《互联网信息服务算法推荐管理规定》	国家网信办等四部门	首次对算法推荐服务进行专门规制，强调算法透明度、用户选择权和算法公平性。
2023年1月	《互联网信息服务深度合成管理规定》	国家网信办等三部门	针对"深度伪造"技术，确立了标识、溯源和主体责任等核心监管规则，防范虚假信息风险。
2023年8月	《生成式人工智能服务管理暂行办法》	国家网信办等七部门	中国首部专门规制生成式AI的法规，确立了"发展与安全并重"及分类分级监管原则，是敏捷治理的典范。
2023年10月	《全球人工智能治理倡议》	中华人民共和国	在国际舞台提出中国方案，倡导发展、安全、治理三维度的全球共治，强调发展中国家权益。
2025年8月	《关于深入实施"人工智能+"行动的意见》	国务院	提出以"人工智能+"赋能千行百业的国家行动纲领，明确了强化政策法规保障、推进健康发展的要求。
2025年9月	《人工智能安全治理框架》2.0	全国网络安全标准化技术委员会	系统性地提出了基于风险分类（内生、应用、衍生）的敏捷治理和分级管理方法论，是未来统一立法的重要理论基础。
2025年9月	《人工智能生成合成内容标识办法》	国家网信办等多部门	细化并强制要求对AIGC进行显式或隐式标识，旨在建立内容的可追溯管理体系。

002 > 司法实践中的规则探索：典型案例的启示

在正式、统一的《人工智能法》尚未出台的背景下，中国的司法系统，特别是以三个互联网法院为代表的专业法庭，正扮演着"事实上的规则塑造者"的角色。通过对一系列前沿、疑难案件的审理，法院不仅在具体个案中定分止争，更重要的是，它们在能动地解释现有法律，将其适用范围延伸至人工智能带来的新场景，从而在实践中探索并确立了一系列重要的裁判规则。这些司法判例如同探路石，为未来的立法提供了宝贵的实践经验和理论素材，形成了一条司法与立法之间的动态反馈回路。

2.1 AI生成内容的权利归属："过程性智力投入"的灵活探索

人工智能生成内容（AIGC）的著作权归属，是全球范围内悬而未决的法律难题。北京互联网法院审理的"AI文生图"第一案（李某某诉刘某某案）为此提供了"中国答案"。在该案中，法院首次明确，利用人工智能生成的内容，如果能够体现出人类用户的"独创性智力投入"，就应当被认定为作品，受著作权法保护。

当然，"智力投入"标准也有一定限度。上海"提示词"案和苏州"蝴蝶椅子案"侧面证明了该标准的有限性和对称性。

在苏州"蝴蝶椅子案"（全国首例否认AI文生图可版权性的案件）中，法院否定了AI生成图片的独创性。理由是，原告输入的提示词"属于相对简单的叠加"，"对画面元素、布局构图等描述缺少差异性"，且被告举证证明在原告之前已有类似概念的作品出现。

同样，在上海"提示词"案中，法院也认定提示词"缺乏作者的个性化特征""属该领域常规表达"。

将上述三地法院的判决合并分析，可以清晰地看到我国的司法裁判，已经为"智力投入"标准建立了一个相对宽松的"标尺"：

- **当人类的"智力体现"独特、个性化、投入高时**，其作品（或其贡献）能跨越了"思想"的门槛，构成受保护的"表达"。
- **当人类的"智力体现"常规、简单、缺乏差异性时**，其贡献则停留在"思想"层面，不受保护。

上述判决尤其是北京判例，引起了各界的讨论，争议最大的部分在于法院的认定是否已经突破了现有著作权法设定的保护框架。

2.2 人格权的延伸保护：AI合成声音与虚拟形象的法律边界

生成式AI的发展，使得对个人声音、肖像乃至整体人格形象的模拟达到了前所未有的逼真程度，由此引发了新型的人格权侵权风险。对此，司法实践展现出了敏捷的适应性和解释力，将传统人格权保护的边界拓展至这些虚拟领域。

在"AI声音侵权案"（殷某某诉某智能科技公司案）和"AI名人声音带货案"（李某某诉某文化传媒公司案）中，法院确立了一个核心的认定标准——"可识别性"。法院认为，无论是通过AI技术合成的声音，还是对录音制品的AI化处理，只要其最终效果能够让一般社会公众或相关领域的听众，根据其音色、语调、发音风格等特征，与特定的自然人建立起清晰的对应关系，那么这种AI生成的声音就落入了该自然人"声音权益"的保护范围。未经本人许可，将这种具有高度可识别性的AI声音用于商业用途（如文本转语音产品、直播带货），就构成了对其人格权的侵害。

更进一步，在"AI陪伴者"案（何某诉某人工智能科技有限公司案）中，法院将保护范围从单一的声音或肖像，延伸到了一个更为综合的"虚拟形象"。在该案中，被告的软件允许用户上传公众人物何某的姓名、肖像，并与其他用户共同"调教"AI，为其注入特定的性格、语言风格和互动模式，从而创造出一个高度拟人化的AI虚拟角色。法院认定，这种行为已经超越了对单一肖像或姓名的使用，而是对何某人格特征的综合性利用，形成了一个与其本人高度关联的虚拟人格。未经许可创设并使用这种虚拟形象，不仅侵犯了其姓名权和肖像权，更因其可能对个人形象和声誉造成歪曲，侵害了由人格尊严和人格自由所构成的一般人格权。

这些判决清晰地表明，中国司法界正在构建一个立体的、多层次的人格权保护体系，确保自然人的人格权益不会因为AI技术的虚拟化、数据化而被削弱。其核心法理在于，若AI的产出物在客观效果上能够指向一个特定的、可识别的自然人，那么该自然人的人格权就应受到法律的

保护。

2.3 平台责任的再定义：从"避风港"到算法实质参与者的转变

在人工智能时代，网络平台不再仅仅是用户生成内容的被动展示渠道，其复杂的算法设计和运营规则，往往深度介入甚至主导了内容的生成与分发。司法实践敏锐地捕捉到了这一变化，并开始重新审视和界定平台的法律责任，逐步打破了传统的"通知-删除"避风港原则的适用边界。

在上述"AI陪伴者"案中，平台方辩称自己仅提供技术服务，侵权内容由用户上传，应适用避风港原则免责。然而，法院驳回了这一主张。法院深入分析了平台的产品设计和算法机制，发现平台并非中立的技术提供者。相反，它通过设定规则、设计"调教"算法，主动地组织、鼓励、引导用户参与到创设侵权虚拟形象的过程中，并从这种核心功能中直接获益。因此，法院认定平台在侵权内容的生成中扮演了"实质性参与者"和"共同创作者"的角色，应当作为内容服务提供者承担直接的侵权责任。

而在另一起"平台误判AIGC案"（唐某某诉某科技有限公司案）中，平台因其AI检测算法错误地将用户原创内容标记为AI生成并予以处罚，而被判承担违约责任。该案的判决逻辑尤为关键：法院认为，平台作为算法的掌控者和决策的作出者，对其自动化决策的结果负有"适度的解释说明义务"。当用户对算法的判定提出异议时，平台不能简单地以"算法结果"为由推卸责任，而应就其判断依据和决策逻辑提供合理的解释。由于平台未能做到这一点，其处罚行为缺乏事实依据，构成违约。

这两个案例共同揭示了一个重要的司法趋势：法院正在穿透平台"技术中立"的表象，对其算法在内容生态中的实际作用进行实质性审查。如果平台的算法本身深度参与、组织或引导了侵权行为的发生，其法律责任将从间接责任升级为直接责任。同时，平台对其算法决策的透明度和可解释性，也正在成为一项重要的法律义务。这种司法导向，无疑对平台企业提出了更高的合规要求，促使其在追求技术效率的同时，必须将法律责任和伦理考量嵌入到算法设计的核心之中。

2.4 合理使用的拓展：以分类施策和包容审慎的态度界定法律责任

杭州互联网法院在"上海某文化发展有限公司诉杭州水母智能科技有限公司著作权侵权及不正当竞争案"（以下简称"**奥特曼案**"）中，提出对生成式人工智能服务的侵权认定采取分类分层的策略：

- 数据输入和数据训练阶段：这一阶段主要目的是学习、分析在先作品所表达的思想感情、语言特征、特色风格等内容，从中提取规则和模式，以便后续进行转换性创作新作品，宜采取相对宽松包容的认定标准。
- 内容输出和内容使用阶段：该阶段直接面向公众，涉及侵权内容的传播和扩散，宜采取相对从严的认定标准。

关于AI训练阶段的作品使用问题，法院认定：生成式人工智能的创设与发展需要在输入端引入巨量的训练数据，不可避免会使用他人作品。在数据训练阶段，如果使用他人作品的目的并非再现作品的独创性表达，且未影响权利作品正常使用或不合理地损害相关著作权人的合法利益，则可以被认为是**合理使用**。因此，本案也宣告了中国司法实践对著作权合理使用制度的"扩容"。稍显遗憾的是，本案由于是用户上传数据，对于训练数据"有毒性"是否导致合理使用原则无法适用等关键问题，仍待后续探讨。

综上所述，这些司法判例共同构成了一部动态演进的"AI习惯法"，它们在具体场景中为AIGC的权利归属、人格权的保护边界、训练数据版权保护的责任划分以及平台责任的归属等提供了极具价值的裁判指引。这种司法能动主义，不仅有效填补了现行法律的空白，更重要的是，它通过个案裁判的方式，对前沿法律问题进行了"压力测试"，其所确立的法律概念，极有可能被未来的《人工智能法》吸收和采纳，从而完成从司法实践到成文立法的转化。

然而，随着人工智能技术的系统性影响日益显现，上述敏捷但分散的治理模式也开始暴露出其固有的局限性。当前，推动一部统一、综合的《人工智能法》或许才是长远之策。只有在"法律"这一高位阶层面制定综合性的人工智能法律制度，才能有效统筹协调各方利益，确立国家层面统一的人工智能治理体系，发挥法治在人工智能时代的根本性、稳定性与长远性作用。

基于对中国AI治理演进脉络和司法实践的深入理解，本章将聚焦于六大核心议题，系统分析它们在未来统一的《人工智能法》及相关配套法规中可能呈现的制度设计与核心考量。这六大议题——支持研发、建设基础设施、完善伦理、监测风险、创新监管、促进健康发展——可视为中国"发展与安全并重"治理理念的若干切入点。

3.1 支持人工智能基础理论研究和算法等关键技术研发

政策目标：中国的国家战略始终将实现高水平科技自立自强置于核心位置。《新一代人工智能发展规划》和最新的《关于深入实施"人工智能+"行动的意见》均反复强调，要加强人工智能基础理论研究，加速推动"从0到1"的重大科学发现，并支持基础模型、关键算法等核心技术的自主创新与突破。立法的首要任务之一，便是为实现这一宏大目标提供制度保障和激励。

可能性分析：未来的立法框架将可能采取一种双管齐下的策略。

一方面，**构建安全可控的开源治理体系**。《框架2.0》明确提出，要在"培育发展开源创新生态的同时，同步提升开源生态安全能力"。这预示着未来的法律将对开源活动进行规范，具体措施可能包括：

1.**明确开源提供方的责任**：要求开源基础模型的提供者履行必要的风险告知义务，如在发布时附带详细的技术文档，说明模型的已知缺陷、潜在偏见、安全漏洞和适用范围。

2.**界定"禁止性"使用行为**：法律可能会授权或鼓励开源社区和提供方，通过开源协议明确禁止将模型用于非法目的，例如制造虚假信息、进行网络攻击或开发违禁武器等。

3.**赋予一定的责任豁免**：针对开源模型的责任豁免设计将成为现实考量。但这种"免责"并非绝对，而是有明确的边界。开源模型开发者若要享受类似于"安全港"的保护，至少需要履行基础的透明度义务和风险防范义务，例如提供说明文档、采取基础技术措施限制生成违法信息。

另一方面，**强化并细化知识产权保护规则**。如上所述，各地互联网法院已经有司法实践示例，笔者建议，未来的《人工智能法》或配套的知识产权法规，可以将这一司法原则上升为明确的法律规则。这可能包括：

1.**保护训练数据中的知识产权**：明确在模型训练阶段使用受版权保护数据或数据权益的法律边界，探索建立合理的使用许可或补偿机制，以回应数据权利方的关切。

2.**分步明确人机作品的权利归属原则**：以"实质性贡献"为核心标准，确立在开发者、使用者、数据提供者等多方参与的复杂场景下，AIGC权利的归属与分配规则。

这种鼓励与开放的立法设计，有助于创造一个既能激励个体或企业进行颠覆式创新，又能确保开源生态整体健康、安全、开放的良性发展环境。

3.2 推进人工智能基础设施建设

为实现综合性治理目标，必须为人工智能的两大基石——"数据"和"算力"——提供统一、安全、高效的法律基础设施。《"人工智能+"行动意见》已将"强化智能算力统筹"和"加强数据供给创新"作为核心基础支撑能力来部署。推进基础设施建设，不仅是技术和投资问题，更需要坚实的法律框架来解决数据产权、数据流通、算力调度和网络安全等一系列复杂问题。

在算力层面，立法可为"全国一体化算力网"的建设提供政策和法律依据。这可能包括：

1.**制定算力基础设施安全标准**：法律将明确国家级智算中心、超算集群等关键信息基础设施的安全防护等级和运营要求，确保算力资源的稳定可靠，防范网络攻击和恶意消耗。

2.**规范算力资源的调度与共享**：通过立法确立跨区域、跨主体算力

资源的互联互通标准和调度规则，推动"东数西算"等国家工程的有效落地。同时，鼓励发展标准化、普惠化的算力服务，降低中小企业使用AI的门槛。

在数据层面，立法将致力于构建一个安全、高效的数据要素市场。

1.**完善数据产权与流通规则**：在现有《数据安全法》《个人信息保护法》的基础上，进一步明确不同类型数据（个人信息、工业数据、公共数据）在AI场景下的权属界定、使用边界和收益分配机制。特别是针对模型训练中大量使用网络公开数据的情况，立法需要对自动化采集数据的合规边界进行更清晰的界定。

2.**推动高质量公共数据开放**：法律将推动建立公共数据有条件、合规开放的制度，鼓励政府和公共机构在脱敏处理后，向社会开放高质量的科学、政务等数据集，以支持基础科研和模型训练，这与《"人工智能+"行动意见》中的要求一脉相承。

3.**培育数据服务产业**：立法将支持和规范数据标注、数据清洗、数据合成等新兴数据服务业态的发展，为人工智能产业提供高质量的"数据燃料"，同时确保数据处理全流程的合规性与安全性。

通过为算力和数据这两大基础设施的建设和运营提供明确的法律保障，中国旨在为整个人工智能产业的腾飞铺设坚实、安全的轨道。

3.3 完善人工智能伦理规范

笔者预计，未来立法将会推动人工智能伦理从"软性倡议"走向"硬性约束"。从早期的《新一代人工智能伦理规范》到《"人工智能+"行动意见》中"完善人工智能法律法规、伦理准则"的明确要求可见，伦理治理的制度化、法治化已成为国家层面的共识。

笔者理解，未来立法不排除将通过"制度嵌入"和"流程强制"的方式，实现AI伦理的"硬落地"。

1.**伦理审查制度**：现行的《科技伦理审查办法（试行）》已经要求从事特定AI科技活动的单位设立科技伦理（审查）委员会。未来的《人工智能法》极有可能将这一要求普遍化和强制化，特别是对于那些涉及生命健康、公共安全、司法执法、金融保险等高风险领域的AI开发与应用。

2.**"价值对齐"的法律化**：正如《框架2.0》中反复强调的"价值观对

齐”（Value Alignment）概念。未来立法可能会要求AI系统的开发者和提供者，采取技术和管理措施，确保其产品和服务的设计、训练数据和输出结果，符合中国的法律法规、社会公德和伦理要求，有效规避产生民族、信仰、性别等歧视性内容的风险。而这项义务的履行情况，亦将可能被纳入算法备案和安全评估的审查范围。

3.**保障弱势群体权益**：立法可能会特别关注人工智能对未成年人、老年人、残障人士等群体的影响，要求相关产品在功能设计和服务模式上充分考虑其可用性、安全性和特殊需求，防止“智能鸿沟”的扩大，这在《生成式人工智能服务管理暂行办法》中已有体现。

通过将上述伦理要求，以强制性规范的形式嵌入到AI产品的全生命周期管理中，中国旨在构建一道坚实的伦理“防火墙”，确保技术的发展始终服务于社会福祉。

3.4 加强安全风险监测评估

人工智能风险具有复杂性、突发性和传导性，如何设计一套既能全面覆盖各类风险，又具有可操作性、能够适应技术快速变化的分类分级评估体系，是未来立法的关键。

考虑到未来立法的框架性，笔者认为模型分类监管不会过于复杂，强监管措施将主要适用于具有高系统性风险的AI系统。这包括参数规模巨大、用户数量众多、具备社会动员能力的通用大模型，也可能包括应用于金融、医疗、自动驾驶等关键基础设施领域的专用模型。

表2：《人工智能安全治理框架2.0》中基于风险的应对原则提出的风险分类，为未来立法提供了重要的理论基础

风险类别	具体风险示例	技术应对措施示例	综合治理措施示例
技术内生安全风险	模型偏见与歧视	改进模型架构，扩充训练数据多样性；引入人类监督机制	构建人工智能科技伦理准则；开展伦理审查
	数据投毒与污染	使用来源合法的训练数据；对数据进行严格筛选和清洗	完善数据安全和个人信息保护规范；制定数据标注安全规范
	模型缺陷扩散（开源）	加强基础模型、开源模型安全缺陷传导评估	强化开源生态安全和供应链安全；明确开源提供方责任
技术应用安全风险	输出违法有害信息	建立安全护栏，对输入输出进行动态过滤	推广AIGC可追溯管理（内容标识）；建立投诉举报机制
	被用于网络攻击滥用	提高系统透明度；完善冗余设计与容灾机制	共享人工智能安全风险威胁信息；建设漏洞信息库
	影响关键基础设施运行	设置能力边界；建立"熔断""一键管控"等应急措施	实施应用分类及安全风险分级管理；对高风险应用进行登记备案
应用衍生安全风险	冲击劳动就业结构	（间接）推动AI赋能传统岗位，开展技能培训	加强AI应用就业风险评估；引导创新资源向创造就业潜力大的方向倾斜
	加剧社会偏见与歧视	在算法设计中采取措施规避歧视风险；提升可解释性	建立健全人工智能安全法律法规；加强社会监督
	"自我意识"觉醒失控	确保人类最终控制；设置安全终止开关；预留人工干预窗口	增进协同应对失控风险的共识；加强最终用途管理

3.5 创新性监管

创新监管并非易事，监管工具必须具备技术敏感性和前瞻性，能够有效应对模型黑箱、算法迭代快等新挑战。同时，监管需要从单一的政府主导，转向政府、行业、社会多方参与的协同治理。

笔者理解，未来的安全监管将可能围绕以下几个制度工具展开：

1.**深化和扩展算法备案制度：**目前已对推荐算法和生成式AI实施的备案制度，将被确立为一项基础性的监管工具并予以深化。不排除未来的备案要求将更加详尽，使监管机构能够提前掌握高风险算法的基底，实现"事前"监管。

2.**强制性内容标识与可追溯性管理：**2025年9月施行的《人工智能生成合成内容标识办法》将内容标识从"行业倡议"提升为"法定义务"。法律将强制要求所有AIGC（包括文本、图片、音视频）都必须附加显式或隐式标识，确保其来源可追溯。这一制度是应对虚假信息、保护知识产权和维护内容生态健康的关键技术监管手段。

3.**构建全生命周期安全管理链条：**法律将可能明确AI价值链上不同主体的安全责任。从**模型算法研发者**（需确保模型内生安全、进行充分测试）、**服务提供者**（需建立安全管理机制、履行内容审核和用户保护义务），到**系统使用者**（需遵守法律和协议、不得滥用技术），法律将构建一个完整的责任闭环，确保每个环节都有明确的责任人，实现全过程治理。

4.**建立多方协同的治理机制：**立法将可能参照数据治理模式，鼓励和规范行业协会制定高于法律底线要求的实践准则，支持第三方专业机构开展有关评测与认证，并建立面向公众的风险隐患举报受理机制。这将形成一个政府监管、行业自律、社会监督、用户参与的多元共治格局，提升治理体系的弹性和有效性。

3.6 促进人工智能健康发展

笔者认为，促进人工智能健康发展是未来立法和治理活动的最终目标，即确保监管在有效防范风险的同时，能够最大程度地释放人工智能作为"新质生产力"核心引擎的巨大潜力，服务于经济高质量发展和社会福祉提升，最终实现《"人工智能+"行动意见》所描绘的智能经济和智

能社会新形态。

这里的挑战是，避免因过度监管或"一刀切"式的规定而扼杀创新活力，确保法律框架具有足够的灵活性和前瞻性，能够为新技术、新模式、新业态的发展留出空间。

因此，笔者理解未来的《人工智能法》将不仅仅是一部"管理法"，更将是一部"促进法"，具体表现为以下几点：

1.**设立"监管沙盒"与试点示范制度：**为了鼓励创新，法律将极有可能采纳《框架2.0》中提出的"包容审慎"原则，或将设立"监管沙盒"或安全可控的试点区域，允许创新企业在有限的范围和可控的风险下，测试其前沿技术和商业模式，并给予一定的容错纠错空间。

2.**加强人才培养与国际合作的法律保障：**法律将与国家的人才战略和外交战略相衔接。一方面，为加强人工智能安全设计、开发、治理等领域的人才培养体系提供法律支持。另一方面，可以考虑将《全球人工智能治理倡议》等国际合作主张的原则融入国内法，如推动技术普惠、支持开源共享、增强发展中国家在全球治理中的发言权等，从而使国内立法成为践行中国全球治理理念的载体。

3.**保障应用落地与产业赋能：**未来立法将为"人工智能+"行动的深入实施提供保障。例如，通过制定重要行业领域（如能源、金融、交通、医疗）的大模型/智能体应用安全指南，为AI技术在这些关键领域的安全、有效落地提供清晰的路径图，从而安全地释放行业应用潜力。

综上，这六大重点议题将勾勒出中国未来人工智能立法的核心部分。笔者理解，它将是一部精巧平衡、多目标驱动的法律体系：既有严格的风险管控底线，又有灵活的创新激励机制；既强调自主可控，又秉持开放合作；既立足于解决国内紧迫的治理难题，又怀抱着塑造领先治理规则的雄心。

最终，中国正在构建的，是一个以根本大法为引领，以专项、垂直法规和规范性文件为支撑，以能动司法为补充，深度融合了产业政策、安全监管和伦理引导的复合型治理体系。这一体系的最终成效，不仅将深刻影响中国未来数十年数字经济和社会的发展轨迹，也将在全球人工智能治理的未来图景中，留下浓墨重彩的一笔。

当然，法律的生命在于执行——如何在实践中真正维系好创新与安全之间那道精妙的平衡，才是挑战的开始。

蔡鹏
合伙人
知识产权部
北京办公室
+86 10 5087 2786
caipeng@zhonglun.com

作 者 / 陈际红　陈煜烺　李佳笑

我国生成式人工智能的监管框架及主要法律风险识别

生成式人工智能（Generative AI，以下简称“Gen AI”）正处于快速发展阶段，技术发展提升了生成内容的质量和多样性，推动了Gen AI在多领域的广泛应用，例如自动化文案生成、图像修复与增强、合成语音和药物发现等领域。随着企业陆续推出基于Gen AI的产品和服务，如国产大模型Kimi、OpenAI的GPT-4o API，全球Gen AI的商业化步伐显著加快。技术的发展和商业化的深入也引发了一系列的法律问题、监管问题和伦理问题，主要国家正在加紧制定相应的法律法规，确保技术发展符合伦理规范，防止滥用，如欧盟《人工智能法案》于2024年8月1日正式生效。

中国采取敏捷治理、小切口立法的路径，迅速回应人工智能技术带来的监管、法律和伦理挑战。全球首部专门针对Gen AI治理的法规《生成式人工智能服务管理暂行办法》（以下简称《暂行办法》）于2023年8月15日正式施行，与《科技伦理审查办法（试行）》等法规共同构建了我国Gen AI治理的初步法律框架。2024年以来，《网络安全技术生成式人工智能服务安全基本要求（征求意见稿）》《人工智能生成合成内容标识办法》等文件接踵而至，各地网信部门亦陆续公布大模型备案信息，标志了我国Gen AI治理迈入纵深推进阶段。

（一）大模型备案

《暂行办法》对Gen AI服务采取了包容审慎和分类分级监管的基本思路，以Gen AI服务提供者为主要监管抓手，规定了算法合规、内容合规、知识产权合规、训练语料合规、数据标注合规等一系列的法定义务。

尤其是，《暂行办法》要求对于具有舆论属性或社会动员能力的、直接面向境内公众提供的Gen AI服务，应开展Gen AI服务的安全评估和备案,即"大模型备案"。需备案的服务包括但不限于具备文字生成、图片生成、音频生成、视频生成等功能的Gen AI服务，不具有舆论属性或者社会动员能力的Gen AI服务无需备案。若服务未面向境内公众提供，则不适用《暂行办法》。对于通过API接口或其他方式直接调用已备案大模型能力的Gen AI应用或功能，网信部门要求采用登记方式，允许其上线提供服务。区别于"生成合成（深度合成）类算法备案"，"大模型备案"在流程和侧重点上与之有明显区别。

（二）算法监管

Gen AI技术的底层逻辑是算法和模型，自2021年起，主管部门相继出台了数部关于算法监管的规定。2021年9月17日，国家互联网信息办公室等九部门发布《关于加强互联网信息服务算法综合治理的指导意见》；2021年12月31日，国家互联网信息办公室等四部门发布《互联网信息服务算法推荐管理规定》；2022年11月25日，《互联网信息服务深度合成管理规定》正式出台；2023年7月，《暂行办法》正式发布，其中亦对Gen AI涉及的算法提出合规要求。至此，我国涉及Gen AI算法监管的法律框架正式形成，Gen AI技术开发者及服务提供者应当依法履行算法相关监管要求，比如进行相关算法的备案。

（三）互联网信息服务及信息内容监管

基于Gen AI之"内容输入"和"内容生成"的运行模式，在我国，通过互联网向公众提供Gen AI服务一般会构成"提供互联网信息服务"[1]，并需承

1.根据《互联网信息服务管理办法》第2条，互联网信息服务是指通过互联网向上网用户提供信息的服务活动。

担配合相应信息内容监管的义务。具体而言，根据《暂行办法》，Gen AI服务提供者应承担内容生产者责任。

主管部门陆续出台了《互联网信息服务管理办法》《互联网文化管理暂行规定》《互联网视听节目服务管理规定》《互联网新闻信息服务管理规定》《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》等针对互联网信息服务的规定，以及《网络信息内容生态治理规定》等专门针对信息内容治理的规定。除遵循一般性的互联网信息服务及信息内容监管规定外，Gen AI服务提供者应结合自身业务模式（例如是否利用Gen AI技术从事"经营性互联网文化活动""互联网视听节目服务"或"互联网新闻信息服务"等），判断是否需遵循特殊监管要求。

（四）增值电信监管

目前，为了给用户提供更好的产品体验，不少Gen AI服务提供者将Gen AI技术嵌入其他垂直领域进行应用。在我国，基于不同网络产品/服务的具体业态（例如是否涉及信息服务业务，是否涉及交易处理业务等），Gen AI服务可能涉及开展增值电信业务的范畴，进而需遵循《中华人民共和国电信条例》《电信业务经营许可管理办法》等规定，并需参照《电信业务分类目录（2015年版）》（2019年修订）依法取得相应增值电信许可证，常见包括ICP证（即前述互联网信息服务）、SP证、EDI证、IDC证、ISP证等。Gen AI服务提供者应结合Gen AI技术所嵌入的具体应用类型，判断是否需取得相应增值电信许可证。

（五）科技伦理审查

《科技伦理审查办法（试行）》于2023年12月1日起施行，其以《科学技术进步法》等作为上位法，突出了Gen AI等技术的科技伦理因素。若从事人工智能等科技活动组织的研究内容涉及科技伦理敏感领域，该组织应设立科技伦理（审查）委员会，并依法开展科技伦理风险评估和审查工作。此外，《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》亦提出了建立科技伦理审查管理制度并采取技术措施的要求。

Gen AI全生命周期所牵涉的数据合规问题复杂，主要阶段包括模型训练阶段、应用运行阶段和模型优化阶段，通常还会涉及Gen AI开发者、服务提供者、服务使用者等多方主体。除了专门规制Gen AI的相关规定外，Gen AI的参与方在进行数据处理活动中，还要遵循包括《网络安全法》《数据安全法》《个人信息保护法》《网络数据安全管理条例》等在内的基础性数据安全法律法规，如涉及数据跨境传输的，还要符合数据跨境传输的相关规定。

（一）模型训练阶段

大模型的成熟度及生成内容的质量都与训练数据高度相关，故模型训练阶段涉及到大量数据的收集以及对此等数据的清洗、分词（以下简称"Token化"），完成后用于模型训练和验证。数据清洗、Token化及模型训练存在内部性，需要重点关注数据质量（真实性、准确性、客观性、多样性[2]）以及Gen AI模型（及所用于训练的数据）的可靠性与稳健性，并根据《暂行办法》第八条制定标注准则、开展数据标注质量评估、抽样核验等[3]。而对于数据收集的风险则需关注数据源合法性，《暂行办法》第七条[4]即要求Gen AI服务提供者"使用具有合法来源的数据和基础模型"。

大模型的主要数据收集形式及合规风险包括：1）采取网络爬虫等形式爬取数据，引发爬虫的合规风险；2）收集已合法公开的公共数据，需要在开放目的范围内以合理的方式处理数据；3）直接面向数据主体收集数据，如收集其个人信息，则需满足个人信息合规的要求；4）面向数据提供方间接收集数据，核心风险在于确保数据源合规，需对该等数据提供方采取合规管理措施；5）合成数据（计算机模拟生成的数据），应主要关注数据质量。

2.《暂行办法》第七条第（四）项规定："采取有效措施提高训练数据质量，增强训练数据的真实性、准确性、客观性、多样性。"
3.《暂行办法》第八条规定："在生成式人工智能技术研发过程中进行数据标注的，提供者应当制定符合本办法要求的清晰、具体、可操作的标注规则；开展数据标注质量评估，抽样核验标注内容的准确性；对标注人员进行必要培训，提升尊法守法意识，监督指导标注人员规范开展标注工作。"
4.《暂行办法》第七条第（一）项规定："使用具有合法来源的数据和基础模型。"

（二）大模型应用运行阶段

本阶段将Gen AI技术投入部署，包括直接提供2C应用、提供2B应用接口（也称MaaS，Model as a Service，“模型即服务”，具体包括“API-标准化服务”和“API-定制化服务”）或私有化部署，即可实现人机交互。本阶段主要的合规风险包括：1）Gen AI服务使用者可能在使用服务时输入个人信息、公司商业秘密、他人享有著作权的作品，进而导致生成内容时可能存在隐私、数据泄露、侵权等风险；2）Gen AI服务提供者可能收集个人信息，且通常具备输出个人信息的能力，可能构成个人信息处理者[5]，进而存在未充分履行个人信息保护合规要求（例如超目的处理、未就数据共享、数据跨境进行充分告知并取得有效同意等）的风险；3）Gen AI服务提供者提供Gen AI服务将面临可靠性与稳健性、透明性与可解释性、准确性与公平性等方面存在欠缺的风险。

（三）大模型优化阶段

本阶段基于人机交互所收集的数据，可能被用于模型的迭代训练。一方面，此等迭代训练过程同样面临模型训练阶段的可靠性与稳健性、透明性与可解释性、准确性与公平性等方面存在欠缺的风险；另一方面，此阶段的外部风险集中在向Gen AI服务使用者提供服务时，未明确就此等模型迭代训练等处理活动事先告知Gen AI服务使用者并取得有效同意。

003 > 知识产权挑战

区别于既定指令的机械执行，“像人类一样思考”的Gen AI实现了从“复制”到“创造”的跨越，给现有创作模式带来了颠覆性的变化，也给现行的知识产权法律体系带来了挑战。

5.参见《GEN AI合规 FAQs（五）| 企业应用Gen AI需关注的数据安全和个人信息保护问题》，2024年11月8日发表于“TMT法律论坛”微信公众号。

（一）使用享有著作权保护的作品开展模型训练是否构成合理使用

大规模和高质量的训练语料供给是大模型成功的基础。由于多数训练语料属于受到现行著作权制度保护的作品，因此对于大模型企业来说，在传统"授权-许可"模式下通过支付费用进而获得全部许可的经济成本过于高昂且耗时漫长，显然也不太现实。在大模型产业快速发展和应用的过程中，围绕创作激励与产业发展的作品使用行为性质也面临诸多争议和讨论，目前已有多个作者、版权方针对Gen AI模型训练过程中使用未经授权作品的行为提起诉讼。例如，2023年，多名美国艺术家对Stability AI LTD在内的三家Gen AI商业应用公司提起版权侵权的集体诉讼；Getty Images也随之在美国针对Stability AI LTD复制其图片用于训练其Stable Diffusion模型的行为提起诉讼。据公开报道，2023年11月，中国的四位绘画创作者将某社交平台诉至法院，诉称该社交平台未经授权使用了创作者的原创作品作为训练数据，从而生成了与原作高度相似的图片，"侵犯了创作者的合法权益"，目前案件仍在审理中。[6]

我国《著作权法》第二十四条规定了12类合理使用的法定情形，直接论证"Gen AI模型训练的作品使用行为构成合理使用"存在一定难度。具体而言，首先，Gen AI的本质是机器学习，且所开发的Gen AI技术一般具有商业目标，较难被认定"为个人学习、研究或者欣赏"；其次，Gen AI作为一种创造性的内容创作系统，并不存在"为介绍、评论或说明"现有作品的前提，且创作过程中难以量化"适当引用"的标准；此外，即使Gen AI研发一定程度上可以被视为"为科学研究"，但"供教学或者科研使用"的目的限制和"少量复制"也一定程度上导致适用困境。除明确列举的法定情形外，《著作权法》第二十四条还规定了"法律、行政法规规定的其他情形"的兜底条款，以留有一定灵活性。《著作权法实施条例》提出了合理使用的"三步检验判断标准"[7]，即同时符合"特定情形下""不影响原作品的正常利用""没有不合理的损害权利人合法权益"的情况下存在被认定为合理使用的可能性，这也为司法快速应对技

6.参见北京互联网法院:《北京互联网法院开庭审理全国首例涉及AI绘画大模型训练著作权侵权案》，2024年6月20日发表于"北京互联网法院"微信公众号，https://mp.weixin.qq.com/s/cyskAz1cASBaNIYQpGpGsA。
7.《中华人民共和国著作权法实施条例》第21条规定："依照著作权法有关规定，使用可以不经著作权人许可的已经发表的作品的，不得影响该作品的正常使用，也不得不合理地损害著作权人的合法利益。"

术发展进行适应性裁判留出了空间。此外，最高人民法院发布的《关于充分发挥知识产权审判职能作用推动社会主义文化大发展大繁荣和促进经济自主协调发展若干问题的意见》第8条规定："在促进技术创新和商业发展确有必要的特殊情形下，考虑作品使用行为的性质和目的、被使用作品的性质、被使用部分的数量和质量、使用对作品潜在市场或价值的影响等因素，如果该使用行为既不与作品的正常使用相冲突，也不至于不合理地损害作者的正当利益，可以认定为合理使用。"这在一定程度上表明，符合一定要素的合理使用在我国存在适用可能性，也在司法层面针对Gen AI模型训练这一问题存在相关规则的适用接口。[8]

在Gen AI兴起的大背景下，对传统著作权合理使用制度与三步检验法产生了革新要求。面对类型化的合理使用制度适用范围较窄，导致无法适应Gen AI产业发展而引发的著作权侵权问题，在进行充分的利益平衡考量的基础上，可以积极考虑将Gen AI模型训练的作品使用行为纳入合理使用规制框架。

（二）Gen AI生成内容是否具有可版权性

针对Gen AI生成内容可版权性的讨论集中于其是否具有"独创性"以及是否为"智力成果"。Gen AI技术的基本逻辑是基于输入内容进行处理并对外输出内容，因此人在其中的参与因素及智力贡献成为了判断可版权性的重要标准。

国际上，2023年2月，美国版权局撤销了含有Gen AI生成图片的漫画《黎明的查莉娅》（Zarya of the Dawn）的原始版权登记。在此案中，美国版权局认为尽管申请人给出的文本提示影响了人工智能生成内容的方向，但因人工智能并非单纯的编辑工具，该生成过程缺乏可预测性，不受申请人控制，故申请人可基于作品文本的作者身份及其对文字、视觉元素的选择、协调和编排，就文本与图像构成整体登记版权，但该版权保护不适用于人工智能生成的每个单个图像。2023年3月，美国版权局发布《版权登记指南：包含人工智能生成材料的作品》

8.参见张伟君：《论大模型训练中使用数据的著作权规制路径》，载《东方法学》2025年第2期。

（Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence），重点强调了只有当作品包含人类创作因素时，该作品才能够受到版权保护（Human Authorship Requirement），拒绝登记仅由机器或纯粹的机械过程而没有人类作者任何创造性投入或干预的情况下随机或自动运行产生的作品。[9]

在中国，除了已经广泛讨论的"威科数据库案"和"Dreamwriter案"外，北京互联网法院近期针对人工智能生成图片著作权侵权纠纷作出的一审判决认为：原告（用户）进行了一定的智力投入，比如设计人物的呈现方式、选择提示词、安排提示词的顺序、设置相关的参数、选定哪个图片符合预期等。涉案图片体现了原告的智力投入，故涉案图片具备了"智力成果"要件；图片的调整修正过程亦体现了原告的审美选择和个性判断，涉案图片具备"独创性"要件。据此判决：1）涉案人工智能生成图片具备"智力成果"与"独创性"要件，应当被认定为作品；2）原告享有涉案图片的著作权。[10]

无论在中国还是美国，对于Gen AI生成内容的可版权性认定思路基本一致：Gen AI生成内容具备独创性且可充分体现人类的智力活动，是Gen AI生成内容成为版权法意义上受保护的客体的前提。但是，关于生成过程中用户对生成物的控制能力、用户智力贡献在生成物中的具体体现及生成过程的可预测性等诸多事实问题，中国和美国显然出现了认知分歧，从而导致对可版权性的观点相左。

（三）Gen AI生成内容权属属于谁

对于Gen AI生成内容的权属，法律并未就此进行明确规定。在上述北京互联网法院审理的案件的判决中，法院认为原告（即用户）是直接根据需要对涉案人工智能模型进行相关设置，并最终选定涉案图片的人，涉案图片是基于原告的智力投入直接产生，且体现出了原告的个性化表达，故原告是涉案图片的作者，享有涉案图片的著作权。[11]

9.参见《他山之石丨美国如何认定Gen　AI的可版权性？》，2023年3月23日发表于"TMT法律论坛"微信公众号，https://mp.weixin.qq.com/s/F0gg5GG4Ce4pjfujYb1d2g.
10.参见北京互联网法院民事判决书，（2023）京0491民初11279号。
11.参见北京互联网法院民事判决书，（2023）京0491民初11279号。

目前Gen AI服务提供者一般是通过协议等方式对Gen AI生成内容的归属作出约定，一般约定相关权益（包括知识产权）归属于Gen AI服务使用者，Gen AI服务提供者获得相应的使用授权。例如，OpenAI在其用户协议中明确，"Open AI将输出内容的所有权利及权益转让给用户。Open AI可能会基于提供和维持服务而进行使用。由于机器学习的特性，基于类似问题可能会产生相同的回复。由其他用户请求和生成的响应不被视为唯一用户的内容。"[12]

（四）Gen AI生成内容是否存在著作权侵权风险

由于Gen AI需要利用现有作品进行模型训练，并通过依赖训练作品形成的算法模型产生Gen AI生成内容，因此，Gen AI生成内容可能会不可避免地携带了训练作品的记忆或痕迹。Gen AI生成内容可能会呈现出训练作品的某些元素、特征、风格等。一般认为，如果Gen AI生成内容与训练作品在表达上构成"实质性相似"且存在"接触"，则可能存在侵权风险。具体而言，如果生成内容可视为训练作品的"复制件"，则可能落入"复制权"乃至"信息网络传播权"的规制范围；如果在保留作品基础表达的前提下形成了具有独创性的新的表达，则可能构成对训练作品"改编权"的侵害。广州互联网法院近期针对Gen AI平台侵权责任作出的一审判决也明确了前述判断思路[13]，由于案涉作品本身知名度较大，且在各大视频网站均可进行访问、查阅及下载，在无相反证据的前提下，该Gen AI平台存在接触该作品的可能性。生成图片部分或完全复制了作品的独创性表达，构成实质性近似，因此侵犯了对原作品的复制权。另外，生成图片部分保留了作品的独创性表达，并在此基础上形成了新的特征，因此侵犯了对原作品的改编权。

除此之外，由于Gen AI生成内容与训练作品的基因脉络一致性，Gen AI生成内容还可能存在风格模仿的问题，如Erin Hanson风格的图画创作、AI孙燕姿的歌曲，也引发了各界对于风格模仿行为的讨论。鉴于版权保护"思想-表达"二分法的基本原则，风格本身并非一种表达形式，

12.参见OpenAI：Terms of Use，https://openai.com/policies/terms-of-use，2025年3月27日。
13.参见广州互联网法院民事判决书，（2024）粤0192民初113号。

无法受《著作权法》保护。但是司法实践中，对于作品哪些部分构成"思想"，哪些部分构成"表达"往往是原被告双方争议的焦点。

Gen AI业界也认识到了潜在侵权对行业发展的困扰，为了消除消费端的侵权担忧，OpenAI、Google、Microsoft、Adobe和Shutterstock等领先的Gen AI业者，率先给出承诺，如果用户遭受第三方的知识产权索赔，其将为用户承担相应的法律责任。

（五）谁对Gen AI生成内容侵权承担责任

《民法典》规定了网络服务提供者责任承担的一般原则，即网络服务提供者无需为用户利用网络服务的侵权行为承担责任，但对于其知道或应当知道的网络用户侵权行为应及时采取必要措施以避免损害扩大。在Gen AI面向终端用户提供服务场景下，生成内容是算法自身基于对Gen AI服务使用者输入内容的理解，通过算法生成的方式完成。尽管Gen AI服务提供者事实上在算法模型训练和优化过程中，会通过数据选择、调参入模等方式对Gen AI生成内容产生影响，但对于最终Gen AI生成内容"选择、编辑、修改"的"输入-输出"这一过程，是由Gen AI服务使用者与算法共同完成的，Gen AI服务提供者本身对此控制较为有限，是否可以据此推定Gen AI服务提供者对生成内容侵权"明知"或"应知"仍有待厘清。

在上述广州互联网法院审理的案件判决中，法院依据《暂行办法》的有关规定认定涉案平台作为Gen AI服务提供者对于侵权行为未尽合理注意义务，从而承担侵权责任。[14]此案的裁判引发了一定的争议，关于Gen AI侵权责任的认定和分配问题仍将会持续引起争论。

陈际红
合伙人
知识产权部
北京办公室
+86 10 5957 2003
chenjihong@zhonglun.com

14.参见广州互联网法院民事判决书，（2024）粤0192民初113号。

作 者 / 蔡鹏

回顾与展望：欧盟人工智能监管趋势探析

2016年3月，由谷歌开发的围棋机器人Alphago在与世界著名围棋选手李世石的对局中获胜，成为第一个战胜围棋世界冠军的机器人。“人工智能”从科技领域的小众词汇，一跃成为现象级的热点话题，高调进入了大众视野。近年来，随着大数据时代的数据洪流和运算能力的指数级跃升，以DeepSeek等为代表的人工智能大模型持续取得突破，人工智能正在快速嵌入社会运行体制机制，广泛应用于社交娱乐、交通物流、金融、医疗健康、教育科研、智能制造、城市管理、环境保护等诸多领域。在重塑人类生产生活范式的同时，人工智能技术创新亦引发了算法“黑箱”、深度伪造、模型偏见、系统安全性不足、科技伦理等诸多新型社会治理风险，亟需通过制定新的制度规范予以回应。

2021年4月，欧盟率先提出《人工智能法》草案，并于2024年7月通过了《人工智能法》这一全球首部综合性人工智能监管法律，对人工智能实施基于风险识别的分级监管。近年来，我国亦在持续推进对于人工智能监管的立法探索，并已出台《生成式人工智能服务管理暂行办法》《人工智能生成合成内容标识办法》等，对呈现爆发式增长的生成式AI服务的监管需求作出回应。然而，时至今日，我国仍未出台任何统一的人工智能监管规则，未来我国对于人工智能将采取何种监管态度、如何构建监管框架、采取哪些具体监管要求，仍然是理论与实务各界热切讨论的重要议题。

他山之石，可以攻玉。欧盟《人工智能法》在立法思路、监管工具和合规水位方面提供了其监管语境下的示范文本，无疑将成为我国未来完善人工智能监管框架的重要参考。基于此，我们通过本文，对欧盟在《人工智能法》框架下的人工智能监管思路的演进历程进行了梳理和分析，为相关企业识别、了解人工智能最新监管趋势提供参考。

001>《人工智能法》：从应用到模型的监管思路演进

早在2015年，欧盟便已表露出对于人工智能的监管关注。2015年1月，欧盟议会法律事务委员会决定成立专门研究机器人和人工智能发展相关法律问题的工作小组，并于2016年5月发布《关于就机器人民事法律规则向欧盟委员会提出立法建议的报告草案》。2017年1月，欧盟议会法律事务委员会就该报告通过了一份决议，建议欧盟委员会就机器人和人工智能提出立法提案，该决议随后于2017年2月获欧盟议会表决通过。至此，欧盟的人工智能立法历程正式拉开了帷幕。

2020年2月，欧盟委员会发布《人工智能白皮书》（White Paper on Artificial Intelligence），描绘了欧盟AI监管框架的基本蓝图：建立一个欧盟范围内的人工智能统一监管框架，并要专注于降低AI技术应用对于基本权利保护、安全以及责任制度有效运作等方面带来的潜在风险。白皮书发布后在全球范围内引发了激烈的立法争论，企业、民间团体甚至不同国家/地区的监管部门之间纷纷"下场"，围绕人工智能的定义、监管范围、监管手段等方面的意见众说纷纭，莫衷一是。2021年4月，在经过多轮意见征集、研究讨论后，欧盟委员会终于正式发布了首版《人工智能法》草案。作为全球首部针对人工智能进行系统规制的法律文本，该版草案聚焦人工智能的应用风险，将应用于不同场景下的人工智能系统划分为"不可接受的风险""高风险""有限风险"和"最小风险"四个等级，并针对不同级别的人工智能系统量身定制了不同的监管措施。

值得注意的是，在这个阶段，欧盟立法者无论是在监管对象（人工智能系统）还是监管逻辑（按照人工智能系统在不同场景的应用风险进行分级监管）上，目光都主要聚焦在人工智能系统这一"应用层"的监管上。各利益相关方对于草案文本的讨论主要关注的是高风险人工智能系统的所涉领域范围及具体应用场景，不断细化高风险人工智能系统的监管框架。2022年11月，ChatGPT"横空出世"，并随即在全球范围内引发了大模型服务的井喷式增长，也使得越来越多的欧盟立法者认识到了一个问题：现有草案所搭建的面向应用层的人工智能系统的监管规范体系，难以简单套用到尚未"分化"出具体用途、应用场景的通用人工智能（General Purpose Artificial Intelligence，GPAI）模型底座本身。在此背景下，如何监管、防范大模型本身存在的潜在风险，成为欧盟立法者无法回避的另一个重要命题。我们对欧盟立法者对于GPAI模型监管

的关键节点梳理如下：

时间	欧盟《人工智能法》立法中关于GPAI模型的关键节点
2021年4月21日	• 欧盟委员会发布《人工智能法案》初始提案，提出基于风险分级的人工智能监管框架，将人工智能系统按特定用途风险划分为不可接受、高、有限和最小共四个等级进行分级规制，并要求高风险AI系统建立技术文档、风险管理及人工监督机制。
2021年5月至2022年5月	• 欧盟各方对于《人工智能法》草案的讨论聚焦于高风险人工智能系统的所涉领域范围及具体应用场景，不断细化高风险人工智能系统的监管框架。
2022年5月13日	• 法国作为欧盟理事会轮值主席国发布第4a条文本，明确通用人工智能系统是指能够执行广泛任务的系统，如理解图像和语音、生成音频和视频、检测模式、回答问题、翻译文本等，将通用人工智能正式纳入《人工智能法》的立法讨论范围。
2022年12月6日	• 欧盟理事会通过关于《人工智能法》提案的共同立场，该版本新增"通用目的人工智能系统"章节，规定若通用AI系统构成高风险AI系统或其组件的，应当遵守高风险AI系统的相关监管规则。
2023年6月14日	• 欧盟议会通过《人工智能法》的谈判授权草案，法案进入欧盟议会、欧盟委员会、欧盟成员国的"三方谈判"阶段。 • 该版本针对通用人工智能的监管逻辑围绕"基础模型（foundation model）"展开，将针对基础模型提供者的合规义务放置在高风险人工智能系统章节中，并对其设置了风险管理、数据治理、透明度、质量管理、防止生成非法内容等专门义务。
2023年6月至2023年12月	• 欧盟范围内围绕通用人工智能的监管开展激烈讨论，主要涉及通用人工智能模型的风险分级框架、透明度与数据治理等方面。 • 各利益相关方的立场分歧：例如，比利时、荷兰等国支持对系统性风险模型实施更严格监管，以强化消费者权益和基本权利保护；法国、德国、意大利等国担忧严格规则阻碍本土AI企业发展。
2023年12月8日	• 欧盟议会、欧盟理事会、欧盟委员会在经过长达36小时的"马拉松式"谈判后就《人工智能法》达成临时协议，同意对"通用人工智能模型"进行分层监管： ■ 所有通用人工智能模型需遵守透明度要求，包括但不限于起草技术文档、遵守欧盟版权法及发布训练内容摘要等。 ■ 具有系统性风险的大模型应遵守更严格的监管要求，包括但不限于开展模型评估、进行对抗性测试、报告安全事件等。
2024年3月13日	• 欧盟议会表决通过《人工智能法》，通过后的法案正式文本会在欧盟公报上公布并在公布之日起20天后生效。
2024年7月12日	• 《人工智能法》最终文本正式公布。正式版在第5章对"通用人工智能模型"进行了专章规定，明确了对于"通用人工智能模型"的分层监管规则，并豁免了开源模型的部分合规义务，仅对其施加最低限度的透明度义务。

并针对不同风险层级的人工智能系统相关条款设置了不同的实施节点。其中，适用范围、定义等一般条款以及被禁止的人工智能实践条款于2025年2月2日首批实施，通用式人工智能监管相关条款被置于2025年8月2日起实施，高风险及有限风险相关规定则被排在2026年8月2日之后实施。

002 > 模型监管：技术创新与风险治理的动态平衡

1.《人工智能法》中的通用人工智能模型监管

《人工智能法》对于GPAI模型同样采取了基于风险的分级监管模式，依照模型的影响能力[1]，从GPAI模型中进一步细分出对于欧盟内部市场具有高影响能力的"具有系统性风险的GPAI模型"，并对其进行特殊监管。具体而言，GPAI模型的提供者须自行评估其模型是否"具有系统性风险"，并在评估认为可能具有系统性风险之日起2周内依法向欧盟委员会申报，欧盟委员会也可自行指定具有系统性风险的GPAI模型。对于具有系统性风险的GPAI模型提供商而言，其需要在遵守下方表格所列的基本合规义务的基础上，进一步履行针对具有系统性风险的GPAI的特殊合规义务：

通用式人工智能模型提供者的基本合规义务	1）编制和更新覆盖该模型的训练、测试过程及其评估结果的技术文件，并应要求成员国主管机构和欧盟人工智能办公室提供； 2）向嵌入该模型的人工智能系统提供者提供信息，使其充分了解该模型的能力和局限性； 3）制定版权政策，确保满足《欧盟单一版权指令》中的文本或数据挖掘的版权例外； 4）按照欧盟人工智能办公室提供的模板，起草并公开关于通用式人工智能模型的训练内容的详细摘要； 5）如通用式人工智能模型提供者为非欧盟企业，其还应当在其模型投放市场前，在欧盟境内指定一名授权代表，代表其履行《人工智能法》下的义务。
具有系统性风险的通用人工智能模型提供者的特殊义务	1）使用标准化工具对模型进行评估和记录，包括开展对抗测试，以识别和降低系统性风险； 2）识别和减轻欧盟层面可能存在的系统性风险； 3）跟踪、记录和及时向欧盟人工智能办公室和成员国主管部门（如需）上报严重事件和可能的纠正措施； 4）确保对该类模型及其物理基础设施提供足够水平的网络安全保护。

1.根据《人工智能法》附件十三的规定，评估模型是否构成"具有系统性风险的通用式人工智能模型"，需考量模型参数数量、数据集的质量或大小、训练模型的计算量、模型的输入和输出模式、模型的自主能力、用户数量等因素进行综合分析。

值得关注的是，欧盟还设置了一条"推定"具有系统性风险的定量标准：如果模型用于训练的累计计算量（以浮点运算（FLOPs）计）大于10^25的，可推定该模型具有高影响能力。实际上，这一推定标准无疑为"具有系统性风险"划定了非常高的门槛，时至今日仍仅GPT-5、LLaMA等行业领先的大模型能达到这一算力水平。这也一定程度上反映了欧盟对于GPAI的监管态度：为具有强大技术、资金能力的大型模型厂商（通常为美国、中国企业）施加较高合规义务，防范、减轻大模型引发的系统性风险，同时避免对欧盟本土的中小型模型厂商带来过高的合规负担，以支持本土人工智能模型厂商的技术创新和业务发展。此外，《人工智能法》还授权欧盟委员会可以修订前述推定的阈值标准，亦佐证了欧盟立法者强调GPAI技术创新与风险治理的动态平衡的监管思路。

2.欧盟通用人工智能行为准则

《通用人工智能行为准则》以下简称**《行为准则》**，Code of Practice是《人工智能法》为GPAI提供商提供的一份过渡性合规方案，旨在解决GPAI模型提供商在《人工智能法》GPAI监管规则生效到欧盟出台统一技术标准这一期间的合规问题。大模型厂商可以通过自愿签署并遵守行为准则来证明其已履行《人工智能法》所规定的义务，未签署该准则的企业则需实施替代性的合规措施来证明其合规性。

2024年9月，欧盟人工智能办公室正式启动了《行为准则》的起草工作，曾先后于2024年11月14日、2024年12月19日、2025年3月11日发布了三版草案，并于2025年7月10日发布了最终版本。2025年8月1日，欧盟委员会与欧盟人工智能委员会确认《行为准则》是GPAI模型提供商证明遵守《人工智能法》的充分工具。《行为准则》主要包括透明度、版权和安全保障三个部分，从启动起草到正式出台，其经历了来自各界利益相关方的激烈博弈，并整体呈现出"概括式规定-细化、严格的具体操作规范-更为精简、灵活的合规落地方案"的趋势：

- **第一版草案：**该版本以《人工智能法》规范为基础，为《行为准则》搭建了包括透明度、版权和系统性风险三大板块的初步框架，并概括性地规定了各板块下的相关合规机制和措施，并未对具体操作细节作

出细化规定。

- **第二版草案：**对第一版草案进行了大幅细化，并增加了模型提供商为遵守《人工智能法》所需满足的"关键绩效指标"（Key Performance Indicators，KPIs）：在透明度方面，细化了模型技术文档所需包含的各类信息；在版权方面，明确了起草并实施内部版权政策、发布内部版权政策摘要、对第三方数据集进行版权合规评估等方面的具体KPI；在系统性风险方面，通过制定详细的KPI详细阐述了系统风险的识别、分析、评估和缓解措施。此外，该版本吸纳了主管部门对于开源模型的监管关注，要求开放权重模型的开发者承担与闭源系统（封闭源代码，仅提供API访问）同等的风险评估义务，打破"开源即免责"的行业惯性。
- **第三版草案及最终版：**精简结构且放弃了第二稿中KPI式表述，整体上取消了第二稿所设置的精细化、量化的合规评估要求，并提供了部分信息文档模板，特别是在针对系统性风险的监管上，为模型提供商提供了更为灵活的合规方案。例如，第二版草案中详细描述了具有系统性风险的GPAI模型的安全缓解措施，第三版草案中则转向提出模型的安全目标，而非提出具体的工具或技术。
- **最终版：**从规则导向转向监管目标导向，删除第三版草案中的"承诺"部分，将其义务内容融入到透明度、版权和安全与保障板块中，按照义务性质、义务对象划分监管模块和文本结构。此外，最终版还强化了版权权利保留、侵权输出防控等方面的要求，并回应了社会各界对于第三版草案过度弱化对于具有系统性风险的GPAI模型的监管的质疑，强化了风险评估义务、建立安全与保障框架等刚性要求。

003 > 总结与展望

回顾欧盟的人工智能监管立法进程，其经历了从聚焦应用层到深耕模型层的演进，从早期仅关注人工智能在不同场景应用的潜在风险管控，到后期转向关注模型本身风险的监管，构建了一套覆盖从模型训练到AI应用的全链条监管体系。而在具体的模型监管措施方面，欧盟始终追求技术创新与风险治理的动态平衡：

- 一是"抓大放小"，在对具有系统性风险的大型模型施加强力监

管义务的同时，仅对中小型模型提出最低限度的透明度等合规要求，在防范化解重大风险的同时确保本土模型厂商的创新发展。

- 二是适当放松对于作为AI技术创新重要推手的开源模型的监管要求，豁免其编制技术文档等部分合规义务，但将开放权重模型排除在前述豁免范围之外，避免因开源社区分布式开发导致出现风险管控敞口和责任真空。

- 三是通过精细化的制度设计以调和AI训练及内容生成与版权保护之间的利益冲突，在《行为准则》中要求模型厂商承诺采取多项措施以确保不存在侵犯在先版权权利的情况，在成员国颁布的《人工智能法》实施法案中亦有相关的细化规则。例如，意大利在2025年9月17日通过的《人工智能法案》中回应了AI创作引发的版权模糊问题，明确规定能体现出人的创造性劳动与智力投入的AI辅助作品可受到版权保护；该法亦明确AI驱动的文本与数据挖掘（text and data mining, TDM）活动原则上仅限于非版权数据，用于经授权的科学研究用途的除外。

蔡鹏
合伙人
知识产权部
北京办公室
+86 10 5087 2786
caipeng@zhonglun.com

作 者 / 斯响俊 朱春天

美国人工智能监管框架

001> 美国人工智能法律框架概述

美国实行三权分立制度，将政府权力分为立法、行政和司法三个独立分支，分别由国会、总统和最高法院行使。同时，美国联邦制又划分了联邦与州的权力边界，确立了"二元立法体系"，即联邦和州各自拥有独立的立法机关和立法权限。联邦层面上，法案需由国会参众两院通过且由美国总统签署生效。州层面上，州立法也需由州议会通过且由州长签署生效。

随着人工智能技术的快速发展，美国的人工智能（"AI"）监管正呈现联邦政策调整、州立法先行以及司法案例快速发展的多层次动态。本文重点评述近期美国联邦层面的国家AI立法与战略、加州AI立法动向，以及人工智能知识产权保护的司法实践。

002> 美国联邦层面人工智能立法与政策概览

美国联邦政府在AI领域的监管策略呈现出一种鲜明的特征——相较于对技术本身进行宽泛的基础性规制，立法者更倾向于针对由AI技术应用所引发的特定、可识别的社会危害进行精准打击。这种"下游应用型"监管模式在近期提出并通过的法案中得到了充分体现。然而，由于总统换届，行政层面则表现出显著的政策波动性，反映出不同执政理念在AI治理路径上的根本分歧。总体而言，美国尚未在联邦层面出台一部系统性的AI法令，而是主要依靠倡导性的AI治理原则，鼓励企业自愿采纳相关政策和技术标准。

（一）立法层面——《*TAKE IT DOWN Act*》[1]

截至目前，美国联邦层面尚未出台一部真正意义上的综合性AI法案，即没有类似欧盟《EU AI Act》那样系统、全面地对人工智能的开发、部署、风险分类、合规义务等进行统一规制的联邦法律。值得注意的是，2025年5月美国国会高票通过并由特朗普总统签署了《*TAKE IT*

1.原文参见：https://www.congress.gov/119/bills/s146/BILLS-119s146es.pdf.

DOWN Act》（全称为《*Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act*》，即《通过阻止网站和网络上的技术性深度伪造来应对已知剥削的工具法案》，下称**《删除法案》**）。《删除法案》以参众两院罕见的跨党派压倒性支持通过，显示出各界对打击AI深度伪造侵害行为的共识。

这是联邦层面首部聚焦AI深度伪造危害的法律。《删除法案》针对非自愿私密影像的深度伪造问题，规定未经同意制作或分享他人私密影像（包括AI生成的私密影像）即构成联邦刑事犯罪，最高可判处2年监禁，若涉及未成年人则最高可处3年监禁。同时，《删除法案》建立了"通知—删除"机制，要求自2026年5月19日起受规制的网络服务平台在接到受害者通知后48小时内删除相关内容。

（二）行政和策略层面——《*Winning the Race: America's AI Action Plan*》[2]

拜登政府曾于2023年发布了第14110号行政令，全称为《*Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*》[3]（《关于安全、可靠、可信赖的人工智能开发和使用》行政令），提出"安全、可靠和值得信赖的AI发展"相关原则。但2025年初新上任的特朗普政府迅速废除了前述行政命令，并于2025年1月23日签署了全新的AI行政令，全称为《*Removing Barriers to American Leadership in Artificial Intelligence*》[4]（《破除人工智能领域阻碍美国领先地位的壁垒》），同时要求相关责任主体于2025年7月前制定并向总统提交一份国家人工智能行动计划[5]。

2025年7月23日，白宫正式发布《赢得竞赛：美国AI行动计划》（全称为《*Winning the Race: America's AI Action Plan*》，下称**《美国AI行动计划》**）。《美国AI行动计划》聚焦于加速本土AI创新、提升美

2.原文参见：https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.
3.原文参见：https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.
4.原文参见：https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.
5.根据《Removing Barriers to American Leadership in Artificial Intelligence》第四条，要求在本行政令发布后的180天内，总统科技助理（APST）、人工智能与加密货币特别顾问及总统国家安全事务助理（APNSA），与总统经济政策助理、总统国内政策助理、管理和预算办公室主任（OMB主任），以及APST和APNSA认为相关的各行政部门及机构负责人协同，制定并向总统提交一份人工智能行动计划。

国国内算力和芯片等基础设施，以及强化国家安全和国际竞争力三大支柱。这一战略标志着联邦政策从侧重监管的路径转向"去监管、促创新"方向，弱化管制审查色彩，更多着力于扶持私营机构开展AI研发、扩大产能和推行有利于技术发展的措施。例如，《美国AI行动计划》指示各机构清理妨碍AI发展的"繁文缛节"，鼓励开放"大模型"权重，扩大芯片制造和数据中心投资，并通过强化出口管制和供应链安全巩固美国在全球AI领域的主导地位。

值得注意的是，《美国AI行动计划》虽然倡导减少联邦层面的监管干预，但并未阻止各州立法。相反，在早前的预算法案《H.R.1 *One Big Beautiful Bill Act*》[6]中试图以联邦法冻结州级AI监管权的提案已被参议院否决。白宫改以行政手段间接影响州政策，如考虑在联邦资金分配时评估各州AI监管"友好度"，甚至要求美国国家标准与技术研究院（National Institute of Standards and Technology，下称"**NIST**"）修改其AI标准以符合新政府的理念。

（三）技术标准层面——《*AI Risk Management Framework*》[7]

2023年1月26日，美国国家标准与技术研究院（NIST）制定并发布了《人工智能风险管理框架》（全称为《*AI Risk Management Framework*》，下称"**AI RMF1.0版**"），其作为倡议性指南帮助企业自愿遵守在AI全生命周期识别和管理风险等原则，促进可信、负责任的AI技术发展与应用。

2024年7月，NIST进一步发布了《*NIST-AI-600-1, Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*》[8]，即专门针对生成式AI的风险管理文件，帮助企业应对生成式AI特有的风险并提供相应行动建议。

总体而言，AI RMF1.0版及相关技术标准为倡导性、非强制性文件，旨在为设计、开发、部署或使用人工智能系统的组织提供自愿性指导，帮助其识别、评估和管理AI相关风险，提升AI系统的可信度，但不具备

6.原文参见：https://www.congress.gov/bill/119th-congress/house-bill/1.
7.原文参见：https://www.nist.gov/itl/ai-risk-management-framework.
8.原文参见：https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf.

法律约束力。

003 > 加州层面：生成式AI监管先行者

在缺乏统一联邦立法且各州皆具有立法权的背景下，各州在人工智能立法方面呈现出高度活跃的态势，其立法路径差异显著。其中，加利福尼亚州（"加州"）凭借经济和科技优势，成为全美人工智能立法最活跃的州之一。加州近期围绕生成式AI领域采取了一系列精细立法，针对AI生命周期的不同环节"逐个击破"，包括训练数据透明度、AI生成内容标识等方面，勾勒出模块化的治理思路。主要的新法案和进展如下：

（一）《AB-2013 *Generative artificial intelligence: training data transparency*》（《生成式AI训练数据透明度法案》，下称《AB 2013法案》）[9]

2024年9月28日，《AB 2013法案》由加州州长签署，并将于2026年1月1日正式生效。《AB 2013法案》要求任何面向加州公众开发、提供生成式AI系统的个人或实体（包括2022年1月1日后发布系统的重大更新）必须在其网站公开所使用训练数据集的相关文档，披露训练数据的使用情况。《AB 2013法案》的此项规定与欧盟《EU AI Act》以及中国《生成式人工智能服务管理暂行办法》的训练数据透明度要求类似，凸显出全球重点法域对于生成式人工智能训练数据透明度的关注和重视。

（二）《SB-942 *California AI Transparency Act*》（《加州AI透明度法案》，下称《SB 942法案》）[10]

2024年9月19日，《SB 942法案》同样由加州州长签署，生效日期同为2026年1月1日。《SB 942法案》要求大型生成式AI提供商（月活用户逾100万）在向加州用户提供AI生成内容时对AI生成内容进行标

9.原文参见：https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB2013.
10.原文参见：https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942.

识，包括隐式标识和显式标识。另外，《SB 942法案》要求大型生成式AI提供商免费提供公开的AI内容检测工具，方便公众或任何第三方校验其提供的内容是否由AI生成，从而构建跨平台的检测生态。

（三）《SB-53 *Transparency in Frontier Artificial Intelligence Act*》（《前沿人工智能透明法案》，下称《SB 53法案》）[11]

2025年9月29日，《SB 53法案》由加州州长签署生效，是美国首部专门针对"前沿人工智能"的州级立法。《SB 53法案》适用于在加州运营、且使用超过10^26次整数或浮点运算的计算能力进行训练的基础模型的开发者，包括OpenAI、Anthropic、Google DeepMind、Meta等公司，要求前沿人工智能开发者公开其研发框架，说明如何采纳标准与最佳实践、如何评估其前沿AI模型可能的安全风险以及采取的防范措施等，同时加州应急服务办公室将建立AI安全事件上报和响应机制。

此前加州曾提出《SB-1047 *Safe and Secure Innovation for Frontier Artificial Intelligence Models Act*》[12]（《前沿人工智能模型安全与保障创新法案》），但最终由于其过于强调前沿人工智能"强制性责任监管"而被否决。此次通过的《SB 53法案》凸显出加州对于AI监管的态度从"强制性责任监管"转向较为温和的"透明度+事件响应"监管思路。

总体而言，加州已率先推出多项针对生成式AI的专项法规，从训练数据透明度、生成式内容标识要求到针对前沿大模型的专门立法，填补了前沿大模型的监管空白，预计将推动美国更多司法管辖区跟进立法。

004 > 人工智能知识产权领域的司法实践

除了立法监管，AI技术的快速发展也在司法领域引发了一系列知识产权纠纷。当前的司法实践主要围绕两大核心问题展开：一是AI生成内容能否获得版权保护，二是使用受版权保护的数据训练AI模型是否构成"合理使用"。

11.原文参见：https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202520260SB53.
12.原文参见：https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB1047.

（一）AI生成内容的版权保护

美国版权局在《版权登记指南：包含人工智能生成材料的作品》[13]中强调，只有当作品包含人类作者身份要求（Human Authorship Requirement）时，该作品才能够受到版权保护。美国版权局在《*Zarya of the Dawn*》等AI作品的准予版权注册通知书[14]中，同样强调单个未经人类再创作的AI图像本身不受版权法保护，但自然人对AI生成图像的"选择、安排和协调"（selection, arrangement, and coordination），如具有足够的创造性，则可能获得版权保护。

（二）AI模型训练数据的合理使用

近期美国法院在Kadrey v. Meta Platforms, Inc.和Bartz v. Anthropic两案中均支持将受版权保护的数据用于AI模型训练可能构成"合理使用"，从而不直接构成版权侵权。两案对于AI模型训练及发展具有里程碑意义，凸显出美国司法实践倾向于在保护版权权利人传统合法权益的前提下，不断探寻AI训练数据使用的合法边界。

a.Kadrey v. Meta Platforms, Inc.[15]

在此案中，Meta被13位知名作家指控Meta从"影子图书馆"下载其受版权保护的书籍，用于训练Llama大语言模型，涉嫌版权侵权。

法院根据"合理使用"四要素进行分析，即（i）用途和性质（包括使用是否具有商业性及是否具有转化性[16]），（ii）受保护作品的性质，（iii）使用量和实质性，以及（iv）对原作品市场或价值的影响。法院认为AI模型训练属于高度转化性使用（即法官认为训练大语言模型与人类阅读的目的迥异，与原作品的使用目的和方式相差较大），且原告未能证明AI训练直接替代原市场，亦未证明潜在许可市场受损，最终认定Meta的复制行为构成合理使用。

b.Bartz v. Anthropic[17]

在此案中，多位美国作家通过集体诉讼指控Anthropic在未获许可

13.原文参见：https://www.govinfo.gov/content/pkg/FR-2023-03-16/pdf/2023-05321.pdf.
14.版权局决定参见：https://www.copyright.gov/docs/zarya-of-the-dawn.pdf.
15.判决参见：https://storage.courtlistener.com/recap/gov.uscourts.cand.415175/gov.uscourts.cand.415175.598.0_1.pdf.
16.转化性使用是指对原作品的使用是否赋予了新的价值、意义或目的，从而与原作品的使用方式或目的有所区别。
17.判决参见：https://s3.documentcloud.org/documents/25982181/authors-v-anthropic-ruling.pdf.

的情况下从互联网盗版书库获取了数百万本受版权保护的书籍，并购买了一些正版书籍，拆除装订后扫描成电子文件，存储在可搜索查询的中央图书馆数据库中，随后从中央图书馆中调取数据用于训练其AI模型，侵犯了作者的著作权。

美国法院在此案中认定：一方面，通过"合理使用"四要素分析，将受版权保护的书籍文本用于训练AI模型本身可视作合理使用，并不直接构成版权侵权。但另一方面，法官认定Anthropic明知且不当地从盗版网站获取了数百万本书籍，这一获取行为本身构成版权侵权。

由于面临潜在的巨额赔偿责任（若集体诉讼最终败诉，Anthropic可能被裁定赔偿数十亿美元），2025年9月初，Anthropic同意支付总额约15亿美元的和解金与众多作者和解，了结此案。

由此可见，美国AI监管目前呈现出联邦和州加强AI立法的态势，同时通过司法实践不断探究和确立AI知识产权纠纷中法律问题的裁判尺度。可以预见，随着AI技术和应用的快速演进，美国各级监管框架也将不断调整完善。在此过程中，立法者、监管者与司法机构之间的互动，将共同塑造AI发展的法律边界和合规要求。出海企业和从业者需要持续关注这些动态，确保在享受AI创新红利的同时，遵守最新的法律规范，降低潜在的法律风险。

斯响俊
合伙人
公司业务部
上海办公室
+86 21 6061 3771
jaysi@zhonglun.com

chapter 02

人工智能与贸易合规｜地缘政治下的博弈

artificial intelligence and trade compliance : geopolitical dynamics

作 者 / 张国勋 代思浓

从封锁到突围：欧美出口管制措施围堵下的中国人工智能发展的挑战与应对

001> 人工智能的战略意义与欧美对华围堵态势

人工智能（Artificial Intelligence，AI）已成为全球科技与产业竞争的制高点，深刻重塑着经济结构、安全格局与治理模式。AI不仅是一项技术，其研发与应用水平，已成为衡量国家科技竞争力、产业现代化水平与制度创新能力的重要标志。

中、美、欧三方均将人工智能视为决定未来国际格局的"战略基石"。中国将AI作为推动高质量发展的核心引擎，强调科技自立自强与安全可控，致力于构建完整的人工智能产业生态和算法创新体系。[1]美国将AI上升为国家安全与经济增长的重要支柱，通过出口管制、政策倾斜等方式保持其技术优势。[2]欧盟则以"以人为本（human-centric）"的理念为导向，试图在伦理治理、隐私保护与国际规则制定中建立规范优势。[3]

然而，随着中国在AI应用和产业化方面的快速崛起，欧美国家开始担忧其在全球科技链与价值链中的相对优势被削弱。美国借"国家安全"与"防止技术外溢"为名，对中国实施芯片、超级计算、AI模型权重及关键算法框架的出口限制；欧盟则在多边机制中推动技术标准和治理规则的"价值化输出"。

这种"硬封锁+软围堵"的组合策略，实质上反映出西方在全球AI竞争中从开放合作转向战略遏制的深层逻辑，即通过管控关键资源与规则制定权，延缓中国人工智能体系的独立化进程，从而维护其长期的科技与制度性优势。

本文旨在梳理和分析欧美在人工智能领域的出口管制的相关措施，探讨其对中国人工智能产业在国家层面和企业层面所产生的影响。

1.参见《国务院关于深入实施"人工智能+"行动的意见》，国发〔2025〕11号，2025年8月21日发布并实施。
2.U.S. Congress, *Regulating Artificial Intelligence: U.S. and International Approaches and Considerations for Congress,* 4 June 2025.
3.European Commission, *"Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act),"* COM(2021) 206 final, 21 April 2021.

（一）美国对人工智能出口管制的相关措施

美国针对人工智能出口管制的主要目的在于防止其领先前沿的人工智能技术，如高性能芯片、超级计算机、AI模型权重等落入潜在对手（如中国、俄罗斯）手中；对内通过推动有利于创新和符合美国利益的人工智能治理模式，从而巩固美国在全球人工智能领域的影响力与主导权。

根据美国《出口管理条例》（*Export Administration Regulations*, 以下简称"**EAR**"），先进人工智能计算能力的出口限制主要通过《商业控制清单》（*Commercial Control List*, 以下简称"**CCL**"）中的类别3（电子）、类别4（计算机）和类别5（电信和信息安全）三大条目来实施。这种双类别框架，使美国能够从硬件、软件到技术实现全链条管控。

2022年10月7日，美国商务部产业与安全局（Bureau of Industry and Security，以下简称"**BIS**"）负责将先进计算芯片、含该芯片的计算机商品及相关"软件"和"技术"纳入CCL，新增3A090适用于特定的高性能集成电路（Integrated Circuit, 以下简称"**IC**"），4A090适用于包含ECCN 3A090所述IC的计算机、"电子组件"和"部件"，以及3B090适用于特定先进半导体制造设备。[4]这三个新的ECCN均因"区域稳定（Regional Stability，以下简称'**RS**'）"理由而被管制，出口、再出口至中国或在中国境内转移时需申请许可证。此外，BIS还在ECCN 5A992和5D992中新增RS许可要求，以涵盖其性能达到或超过ECCN 3A090或4A090标准的物项。[5]BIS还在EAR § 734.9外国直接产品规则（Foreign Direct Product Rules, 以下简称"**FDP Rules**"）中新增第（h）段先进计算外国直接产品规则（Advanced Computing FDP

4.BIS, Implementation of Additional Export Controls: Certain Advanced Computing and Semiconductor Manufacturing Items; Supercomputer and Semiconductor End Use; Entity List Modification, 13 October 2023.
5.BIS, Implementation of Additional Export Controls: Certain Advanced Computing and Semiconductor Manufacturing Items; Supercomputer and Semiconductor End Use; Entity List Modification, 13 October 2023.

Rule）[6]和第（i）段超级计算机外国直接产品规则（Supercomputer FDP Rule）。[7]此规定使得即便芯片在国外制造，但若使用了特定美国技术，也可能被纳入出口许可义务范围。2023年10月25日，BIS修订ECCN 3A090对特定的高性能集成电路的定义与控制，修改相关性能门槛与判定标准，用以防止企业通过降频或更改型号而规避出口管制。[8]2025年1月15日，BIS发布临时最终规则（Interim Final Rule），以加强和完善其在出口管制框架下对最先进AI模型及大规模先进计算集成电路（IC）集群在全球扩散的管控，从而保护美国的国家安全与外交政策利益。[9]具体而言，BIS扩大了对高性能计算芯片的出口管制范围（涵盖ECCN 3A090和4A090项下的集成电路），同时新增了对先进封闭权重（advanced closed-weight）[10]双重用途AI模型的模型权重的出口限制，并为此设立了新的分类编号ECCN 4E091。

除涉及人工智能领域的受控物项出口管制规定之外，BIS以"危害美国国家安全与外交政策利益"为由，将多家中国人工智能及大数据企业和科研机构增列至EAR下的实体清单（Entity List），对中国人工智能产业的发展实施实质性限制。被列入实体清单后，若拟向其出口、再出口或境内转移任何受EAR管辖的物项，均须事先获得BIS许可，且在审查

6.先进计算外国直接产品规则的主要规定：同时满足产品范围和最终用途范围的外国制造物项将落入先进计算外国直接产品规则受到EAR管辖，具体规则如下：
A. 产品范围 （1） 特定的受EAR管辖的软件或技术的直接产品，该软件或技术的ECCN编码为3D001、3D901、3D991、3D992、3D993、3D994、3E001、3E002、3E003、3E901、3E991、3E992、3E993、3E994、4D001、4D090、4D993、4D994、4E001、4E992、4E993、5D001、5D002、5D991、5E001、5E991或5E002；或者相关物项是由位于美国境外的工厂或工厂的主要部分生产，不论该工厂或工厂的主要部分是美国还是外国制造的，该工厂或工厂的主要部分本身属于上述ECCN所列美国原产的特定软件或技术的直接产品；和 （2） 外国制造物项ECCN编码为3A090、3E001（针对3A090）、4A090或4E001（针对4A090）；或者是3A001.z、4A003.z、4A004.z、4A005.z、5A002.z、5A004.z或5A992.z下的集成电路、计算机、电子组件或组件。
B. 目的地和最终用途范围 （1）拟出口至全球任何地区，或将被纳入至非EAR99的任何"零件"、"组件"、"计算机"或"设备"中，且该等物项拟出口至全球任何地区；或（2）由总部位于澳门或D:5组国家，或其最终母公司总部位于上述地区的实体"开发的"技术"，该技术用于掩模（mask）、集成电路晶圆或裸片的"生产"。
7.超级计算机外国直接产品规则的主要规定：同时满足产品规则和最终用途范围的外国制造物项将落入超级计算机外国直接产品规则受到EAR管辖，具体规则如下：
A.产品范围（1）属受EAR管辖并列于ECCN 3D001、3D901、3D991、3D992、3D993、3D994、3E001、3E002、3E003、3E901、3E991、3E992、3E993、3E994、4D001、4D993、4D994、4E001、4E992、4E993、5D001、5D002、5D991、5E001、5E002 或5E991的"技术"或"软件"的"直接产品"；
或（2）是由位于美国境外的工厂或其"主要部分"所生产，而该工厂或"主要部分"（无论产自美国或其他国家）是上述所列ECCN中美国原产"技术"或"软件"的"直接产品"。
B. 国家和最终用途范围
如果申请人"知晓"该外国生产的物项将被：（1）用于中国境内的，或拟出口至上述地区的"超级计算机"的设计、"开发"、"生产"、运行、安装（包括现场安装）、维护（检查）、修理、大修或翻新；或（2）被纳入，或用于"开发"或"生产"将用于中国境内的，或拟出口至上述地区的"超级计算机"的任何"零件"、"组件"或"设备"，则该物项符合此处的国家和最终用途范围。
8.BIS, Implementation of Additional Export Controls: Certain Advanced Computing Items; Supercomputer and Semiconductor End Use; Updates and Corrections, 25 October 2023.
9.BIS, Framework for Artificial Intelligence Diffusion, 15 January 2025.
10."封闭权重"指模型权重未向公众公开，即专有模型（如 GPT-4、Claude、Gemini 等），与开源/开放权重模型（Llama、Mistral）相对："先进"指训练算力 ⩾ 1×10^{26} FLOPs（约等于 GPT-4 量级）的封闭权重模型才被纳入管制；美国《出口管理条例》新增ECCN 4E091，对达到算力门槛的封闭权重参数文件实施出口管控。

中适用"推定拒绝（Presumption of Denial）"的许可政策。[11]该措施导致中国人工智能企业在先进计算芯片、关键算法框架、高精度传感器及其他核心资源的获取方面面临重重障碍，显著制约中国在高性能计算与智能系统研发领域的持续发展能力。

2024年5月22日，美国国会发起加强海外关键出口限制国家框架法案（*Enhancing National Frameworks for Overseas Restriction of Critical Exports Act*，以下简称"**ENFORCE Act**"），该法案旨在修订《2018年出口管制改革法案》（*Export Control Reform Act of 2018*），以防止外国对手利用美国的人工智能和其他关键技术。[12]

2025年1月15日，拜登政府公布《人工智能扩散规则》（AI Diffusion Ruel）该规则已于同年5月13日被特朗普政府撤销。[13]撤销同日，美国政府同步发布三项与人工智能及先进计算芯片出口管制相关指南：（1）提醒业界注意使用中国先进计算集成电路（IC），包括中国制造芯片所带来的风险；[14]（2）警示公众——若允许美国的人工智能芯片被用于中国人工智能模型的训练与推理，可能造成的潜在后果；[15]（3）指导美国企业如何保护供应链，防范被转移或规避管制的行为。[16]

2025年7月23日，美国白宫发布《赢得人工智能竞赛：美国的人工智能行动计划》（*Winning the AI Race: America's AI Action Plan*），阐述特朗普政府确保美国在人工智能领域的领导地位与主导地位的愿景。[17]该计划旨在加速人工智能创新、建设美国人工智能基础设施和引领国际人工智能外交与安全。

（二）欧盟对人工智能出口管制的相关措施

欧盟目前没有像美国那样系统性地通过出口管制措施封锁中国人

11.BIS, Entity List FAQs, Can a listed entity act as purchaser or freight forwarder to transport my shipment of items subject to the EAR to the ultimate consignee or end-user?
12.H.R.8315, Enhancing National Frameworks for Overseas Restriction of Critical Exports Act, 22 May 2024.
13.BIS, Department of Commerce Announces Rescission of Biden-Era Artificial Intelligence Diffusion Rule, Strengthens Chip-Related Export Controls, 13 May 2025.
14.BIS, Guidance on Application of General Prohibition 10 (GP10) to People's Republic of China (PRC) Advanced-Computing Integrated Circuits (ICs), 13 May 2025.
15.BIS, BIS Policy Statement on Controls that May Apply to Advanced Computing Integrated Circuits and Other Commodities Used to Train AI Models, 13 May 2025.
16.BIS, Industry Guidance to Prevent Diversion of Advanced Computing Integrated Circuits, 13 May 2025.
17.White House, White House Unveils America's AI Action Plan, 23 July 2025.

工智能技术，AI芯片或模型，但其通过人权杠杆、多边机制和价值观外交，正在形成一种"软性围堵"。

欧盟对人工智能技术的出口管制主要依托《欧盟两用物项出口管制条例》（*EU Dual-Use Regulation*, Regulation (EU) 2021/821"），该条例于2021年9月生效，是当前最主要的法律依据。[18]2025年9月8日，欧盟委员会更新附件一（Annex I）中的欧盟两用物项出口管制清单。[19]增加与人工智能管制相关的物项为半导体制造与测试设备及材料（例如：原子层沉积设备、外延沉积用设备和材料、光刻设备、极紫外保护膜、掩模与光罩、扫描电子显微镜设备、蚀刻设备）和先进计算集成电路与电子组件（例如：现场可编程逻辑器件及系统）。

值得注意，《欧盟两用物项出口管制条例》在序言中明确指出确保欧盟及其成员国在两用物项领域内充分考虑包括人权在内的相关因素。第5条提及了以人权为导向的物项管控机制，若某物项存在被用于镇压或侵犯人权的风险，即便物项不在Annex I清单中，也必须受出口许可管制。第9条允许成员国出于公共安全或人权理由，对未列入Annex I清单的物项单独设立出口许可要求或直接禁止出口。第10条则要求成员国相互承认彼此的人权导向出口管控。

欧盟在人权导向的技术治理还体现在《欧盟人权与民主行动计划（2020—2027）》（*EU Action Plan on Human Rights and Democracy 2020-2027*）中，该计划的第4部分重点阐述欧盟在数字与人工智能领域的人权外交与治理目标，并提出欧盟将通过多方协作、伦理标准、隐私保护、数据安全与问责机制，推动全球范围内技术发展与人权保障之间的平衡。

《欧盟人工智能法案》（*EU Artificial Intelligence Act*）对人工智能系统在欧盟的开发和使用建立了较为严格的监管机制。[20]通过监管，欧盟将AI系统分为四类风险等级：不可接受风险（Unacceptable Risk）、

18.European Union, Regulation (EU) 2021/821 of the European Parliament and of the Council of 20 May 2021 setting up a Union regime for the control of exports, brokering, technical assistance, transit and transfer of dual-use items (recast), Document 32021R0821.
19.European Commission, 2025 Update of the EU Control List of Dual-Use Items, 8 September 2025.
20.European Union, Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No .167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 an, Document 32024R1689.

高风险（High Risk）、有限风险（Limited Risk）与最低风险（Minimal / No Risk）。[21]虽然该法案本身不是出口管制措施，但其监管系统的等级分类会影响市场准入机制，间接影响欧盟境内的人工智能研发和出口。

003 > 中国的应对路径

（一）完善人工智能方面的出口管制治理体系

面对欧美在人工智能及关键技术领域不断强化的管制与封锁，中国需从制度层面构建以国家核心战略资源保护为导向的科技安全与出口管制治理体系，确保关键要素、核心数据与战略技术的可控与安全。从制度层面完善科技安全与出口管制治理体系，构建既能有效防范外部风险、又能维护开放合作的政策框架。

一方面，应以《对外贸易法》《出口管制法》和《两用物项出口管制条例》为基础，完善战略资源保护机制，形成以风险防控与对等回应为核心的法律框架。通过建立覆盖人工智能、高端芯片、关键算法、稀土材料及核心工业软件的分级分类管控体系，防止国家在关键供应链、底层算力和核心数据环节上形成对外依附或被"卡脖子"的结构性风险。2025年10月9日，中国商务部连续发布四份关于稀土相关物项及技术的出口管制公告。通过对稀土原料、设备及相关技术实施全链条出口管制，反映出国家正以稀土为战略支点，将人工智能纳入国家安全与科技竞争的核心体系。政策旨在防止关键资源与技术外流，限制境外利用中国稀土和技术发展高端制造及潜在军事用途人工智能，同时推动国内产业链自主可控，构建以资源安全、技术管制和人工智能战略相协同的国家科技安全格局。[22]中美两国领导人釜山会晤后达成共识，中方将暂停实施针对10月9日公布的相关出口管制等措施域外管辖部分一年，并将研究细化具体方案。

21.Regulation (EU) 2024/1689, Whereas & Article 36.
22.商务部 海关总署公告2025年第56号 公布对部分稀土设备和原辅料相关物项实施出口管制的决定，2025年10月9日。
商务部 海关总署公告2025年第57号 公布对部分中重稀土相关物项实施出口管制的决定，2025年10月9日。
商务部公告2025第61号 公布对境外相关稀土物项实施出口管制的决定，2025年10月9日。
商务部公告2025第62号 公布对稀土相关技术实施出口管制的决定，2025年10月9日。

另一方面，应加快落实《全球人工智能治理倡议》《人工智能安全治理框架》2.0版，在与国际规则接轨的基础上，强化对国内核心资源与技术标准的自主掌控。推动自主标准与技术规范的制定权与话语权，确保中国在全球人工智能治理规则重塑过程中具备制度影响力。

（二）中美贸易中的物项识别与合规难点

人工智能领域企业通常涉及高性能芯片、超级计算机及AI模型等高技术物项，其产品在研发、生产和出口过程中极易触及多国出口管制体系。为防范合规风险，企业应建立系统的物项归类与技术来源追踪机制，确保在产品设计、制造及软件开发各环节能够准确识别是否涉及美国原产技术或设备。一旦产品性能参数接近EAR下受控物项的技术门槛，企业应主动向美国BIS申请分类咨询或提交合规说明，以降低误判风险，避免被认定为通过"降频"或"更改型号"规避出口管制。

同时，人工智能企业与其他出口经营主体一样，也应对拟出口物项进行系统识别与分类管理。在中国境内出口时，企业需严格对照商务部公布的《两用物项出口管制清单》（以下简称**《清单》**）和《两用物项和技术进出口许可证管理目录》（以下简称**《目录》**），确认产品是否属于管制范围；在涉美出口或再出口情形下，则需依据美国CCL确定其出口管制分类编号（ECCN），判断是否受EAR管辖。

值得注意的是，中国的《清单》和《目录》与美国的CCL虽然在分类体系上均划分为十个大类，但在物项归类逻辑、管制理由判定及技术参数阈值方面存在差异。同一物项在两国清单中可能被归入不同类别，适用的管制理由也不尽相同。这种体系差异容易导致企业在国际贸易中出现重复申报、错位申报或管制义务判断不一致等问题，增加合规成本与出口不确定性。

（三）中欧贸易中的价值导向与合规应对

面对欧盟以人权为核心的"软性管控"模式，中国企业在人工智能出口与对欧合作中，应注重制度合规、透明治理与价值沟通三方面的应对策略。

首先，在企业合规体系建设方面，加入相关人权合规的考量，《欧

盟两用物项出口管制条例》第5、9、10条的所涵盖"人权"考量意味着，即便企业的AI产品未被列入Annex I清单，若存在被用于侵犯人权、大规模监控或内部镇压的风险，仍可能被拒绝出口许可。因此，中国企业在涉欧业务中应建立相应"人权因素"的尽职调查机制，对AI算法、监控系统、数据采集等产品的最终用途与用户背景进行审查，确保符合《欧盟两用物项出口管制条例》等法律法规的标准。

其次，提升数据安全与AI透明度管理水平。企业可以根据《欧盟人工智能法案》的风险等级分类要求，对其产品进行相应的风险归类，并建立相应的合规稳定以便追溯可查，以应对欧盟市场监管要求。

最后，建立与欧盟监管机构的常态化沟通渠道。企业可通过欧盟驻华使馆、行业协会等平台，参与AI治理标准的讨论和行业倡议，及时了解欧盟出口管制与伦理政策的更新动向。对于高新前沿技术出口项目，可主动向欧盟成员国主管机关提交说明性文件，表明产品用途、技术控制措施及人权风险防范机制，以应对合规要求。在此基础上，建议企业加快海外布局，构建"国内国际双循环"相互促进的供应与销售链条。通过在欧盟或关联区域设点布局，推动技术本地化适配与产业链协同，既能够贴近市场响应监管要求，也有助于分散单一市场风险，增强全球竞争韧性。

004 > 结论

总体来看，人工智能已成为全球科技与制度竞争的核心。美国以国家安全为名实施强力监管，通过EAR和实体清单（Entity List）禁止或限制中国获取先进AI芯片与模型；欧盟则以人权与伦理为导向推行软性制度监管，通过《欧盟两用物项出口管制条例》和《欧盟人工智能法案》塑造其全球治理标准。

面对逐步加强监管的国际环境，中国可以从三方面应对。首先，完善出口管制制度，以《对外贸易法》《出口管制法》和《两用物项出口管制条例》为基础，形成安全可控、对等回应的出口管制体系，在制度层面强化关键技术和敏感物项的安全边界，确保国家利益与产业安全的统一。其次，强化企业合规管理。人工智能相关企业应建立完善的物项

识别、技术溯源和客户背景审查机制，确保出口活动符合中外出口管制要求，防范违规出口、技术外溢及被动涉制裁等风险。最后，提升国际参与度与规则影响力。通过积极参与人工智能国际治理和出口管制规则制定，推动全球科技治理体系的多边化与平衡化，增强中国在国际竞争中的制度话语权与战略主动性。

人工智能领域的博弈不仅是技术竞争，更是规则之争。中国唯有在制度完善、创新驱动与开放合作中取得平衡，方能实现人工智能领域的安全与自主。

（戴赟对本文亦有贡献）

张国勋
高级顾问
合规与政府监管部
北京办公室
+86 10 5957 2288
zhangguoxun@zhonglun.com

作 者 / 于治国

美国AI领域贸易与投资限制的制度、实践及应对策略

（一）国家安全驱动的科技管控逻辑

近年来，美国政府逐渐将人工智能（AI）、半导体、量子信息、生物科技等前沿技术视为“战略性技术”，并明确纳入国家安全政策的重要组成部分。其中，美国政府在AI领域的政策取向日益凸显出强烈的“国家安全”导向。AI被认为不仅是未来经济增长的核心动力，更是全球安全与竞争格局的决定性因素[1]，主要体现在以下方面：

1.国家竞争力：AI在自动化、制造、医疗、交通、能源等多个产业的渗透程度越来越高，掌握AI生态的主导权将帮助美国在下一轮产业竞争中取得优势；

2.信息与数据安全：AI技术涉及大规模数据的处理和建模，技术外流可能影响美国对关键信息和个人隐私的掌控能力；

3.经济安全：AI技术外溢可能削弱美国企业的竞争力，导致关键产业链失衡。

此外，美国也在相关政策文件中多次强调国家安全与AI之间的关系。例如：

• 《国家安全战略》（National Security Strategy）（2022）明确提出，AI、先进计算等新兴技术关系到美国在21世纪的战略竞争。

• 《出口管制改革法案》（Export Control Reform Act，ECRA）要求商务部识别并管制包括AI在内的“新兴和基础技术”。

• 《芯片和科学法案》（CHIPS and Science Act）（2022）通过巨额资金支持，强化美国在半导体和AI研发上的优势，同时明确限制其对外转移。

• 《国际紧急经济权力法》（International Emergency Economic Powers Act，IEEPA）授予总统和相关机构法律权力以应对对国家安全构成“非常规威胁”的对外技术转移行为。

1.Framework for Artificial Intelligence Diffusion，
https://www.federalregister.gov/documents/2025/01/15/2025-00636/framework-for-artificial-intelligence-diffusion.
尽管该文件最终被美国政府废止，但其中的考量和政策工具仍然为分析和判断美国AI政策提供了依据。

在这一逻辑下，“保护领先优势、遏制对手发展”成为美国对AI管控政策的重要出发点。具体而言，美国关注的风险点集中在 AI 的若干关键要素：高级算力（GPU、加速卡）、训练数据与数据集、模型权重、核心算法与开发工具链以及相关设计与制造能力。美国不仅通过技术分类来决定出口许可，还强调最终用途、最终用户及潜在风险，形成了一个覆盖出口、投资、科研合作与人才交流的“全链条”管控体系。

（二）政策演进路径

美国对AI的管控政策经历了逐步收紧的过程，体现了从“风险感知”到“主权防护”的转变，呈现出“从点到面、从技术到体系”的演化趋势。

时间阶段	政策特征	主要举措
2018–2020	确立AI为战略技术	将AI列入新兴技术清单；启动实体清单机制；强调军民两用风险。
2021–2023	从安全转向伦理与责任	发布“负责任的AI”框架；推动算法透明、风险评估与国际合作限制。
2024–至今	强调AI技术主权与全面限制	借助对外投资审查制度；限制跨境AI合作与算力输出；将AI与半导体管控联动实施。

这一演进路径反映出美国AI管控政策的三个趋势：（1）从技术出口限制到全链条控制；（2）从风险评估到主权防护；（3）从国内立法到国际协同。美国正在通过立法、行政令和外交合作建立一整套AI安全管控体系，既限制外国企业获取AI核心资源，同时也能防止美国本土技术被间接利用。

002 > 美国主要限制措施概览

（一）AI出口管制体系

美国对AI的出口管制主要依托商务部工业与安全局（Bureau of Industry and Security，BIS）的监管和《出口管理条例》（Export Admin-

istration Regulations，EAR）。其核心措施包括：

1.AI新兴技术清单

2018年起，BIS将AI正式纳入"新兴与基础技术清单"，明确AI包括以下技术内容：

- 深度学习（Deep Learning）、机器学习（Machine Learning）框架及模型；
- 自然语言处理（NLP）技术与大型语言模型；
- 自动驾驶AI系统；
- 图像识别与计算机视觉算法；
- 生成式AI技术（Generative AI）。

此类技术被认为具有潜在的国家安全影响，因此需要通过出口管制等工具和手段进行管控。

2.实体清单（Entity List）

美国通过"实体清单"机制对特定外国企业实施出口管制。部分被列入清单的AI相关公司涉及：

- 提供AI训练芯片、加速卡的硬件企业；
- 从事AI算法研发与大模型训练的公司；
- 提供数据标注、计算平台的云服务商。

被列入清单的公司若要获得美国技术或设备，须申请特别许可，而此类许可通常会被拒绝。该清单更新频繁，相关企业面临高度不确定性。

3.军用最终用途规则（Military End Uses，MEU）

虽然AI被视为民用创新的核心驱动力，但其军民两用特征使得美国在监管上采取"目的地与用途双重审查"原则。根据MEU规则，美国出口商在向外国企业提供AI相关技术或设备时，必须证明其最终用途不会涉及国家安全风险。BIS要求企业建立尽职调查机制，核实最终用户与用途的合规性。

4.外国产品规则（*Foreign Direct Product Rule*，FDPR）

AI领域的FDPR扩展了美国出口管制的适用范围，即便产品在美国境外制造，只要其在设计、开发或制造过程中使用了美国的技术、软件或设备，也受美国管制。

在AI相关产业中，FDPR主要适用于：

• 使用美国EDA（Electronic Design Automation）软件设计的AI芯片；

• 依赖美国云平台训练的大模型；

• 含有美国算法模块的AI应用。

美国通过其出口管制法规规则已经建立起适用不同国别、行业、领域和物项等维度的FDPR[2]。

5.AI扩散框架与尽职调查规则（Framework for Artificial Intelligence Diffusion）

2024年后，美国进一步强化了对AI模型、算法与算力资源的跨境管理[3]：

• 限制AI模型权重和训练数据的跨国转移；

• 要求企业在提供云算力服务前识别客户的国籍与用途；

• 对AI训练平台的共享和转包实施报告义务。

这一系列措施旨在防止外国企业通过云计算或第三方合作规避美国出口管制。

（二）对外投资限制机制

2023年，美国总统签署行政命令，确立了对外投资安全审查制度，并于2025年正式实施。该制度的目标之一是对涉及AI、半导体和微电子等AI相关领域的对外投资进行审查，防止技术通过资本渠道外流、推动竞争对手发展。

1.审查范围

审查机制主要覆盖以下类型的AI投资活动：

• AI模型训练与开发平台；

• AI算力与数据处理基础设施；

• 自动化决策与预测系统；

• 涉及关键算法或模型权重的技术转移。

2.BIS — FDPR / Foreign Direct Product Rules 解读文件，https://www.bis.doc.gov/index.php/documents/compliance-training/3476-foreign-direct-product-rules-update-fdprs-clean-3-19-24-513pm/file.
3.Framework for Artificial Intelligence Diffusion，https://www.federalregister.gov/documents/2025/01/15/2025-00636/framework-for-artificial-intelligence-diffusion.

2.管理模式

类型	内容
禁止类交易	涉及AI在军事、监控、网络安全等高敏感领域的应用；
申报类交易	包括通用AI平台、AI算力服务、数据标注与算法工具；
豁免类交易	如纯财务投资、公开市场股权投资、无控制权的小额投资。

3.处罚机制

对于违反规定的AI相关投资，美国财政部可：

- 处以高达交易金额两倍的罚款；
- 要求强制撤资或剥离资产；
- 对历史交易进行追溯性审查。

美国财政部还要求投资者建立内部合规体系，对潜在的AI投资项目进行技术属性识别与风险评估。

（三）美国CFIUS对AI领域的投资限制

美国外国投资委员会（Committee on Foreign Investment in the United States，CFIUS）对涉及AI的外国投资保持高度警惕。当涉及外国投资者在美投资时，CFIUS会监督和审查有关AI领域的投资情况[4]。[5]

根据其权限范围，在具体项目中涉及AI领域的重点审查对象包括：

- 处理大量个人数据的AI公司（如医疗AI、社交平台AI推荐系统等）；
- 与关键基础设施相关的AI应用（如电网预测性维护、智能交通管理等）；
- 国防或安全敏感应用。

与AI相关的具体监管内容包括：

- 涉及AI在关键基础设施、能源、交通、通信领域的应用；
- 处理或存储敏感个人数据的AI公司；

4.CFIUS ANNUAL REPORT TO CONGRESS，https://home.treasury.gov/system/files/206/2024-CFIUS-Annual-Report.pdf.
5.US compels Saudi fund to exit Altman-backed AI chip startup - Bloomberg News, https://www.reuters.com/technology/us-compels-saudi-fund-exit-altman-backed-ai-chip-startup-bloomberg-news-2023-11-30/.

• 能够通过投资获得AI技术或决策模型访问权限的交易。

（四）AI技术合作与人才交流限制

在AI合作与人才交流方面，美国也有相关的配套限制措施：

1.学术合作

• 限制美国高校与外国高校在深度学习、自然语言处理、生成式AI等领域的联合研究；

• 要求项目资金来源透明，防止"间接技术转让"。

2.人才交流

• 审查外国学者在美参与AI项目的背景，重点关注资金来源及合作对象；

• 部分高敏感AI研究岗位仅对美国公民或长期绿卡持有人开放。

3.技术转让

• 禁止向外国企业转让AI源代码、模型权重和核心训练架构；

• 严格限制跨境共享大规模训练数据集。

4.算力服务

• 禁止美国企业为部分外国客户提供AI训练所需的高性能GPU算力[6]；

• 要求云服务商建立"客户识别机制"，核实AI相关项目的用途与用户背景[7]。

003 > 中国企业面临的主要风险点

美国上述限制措施对中国AI与半导体企业的研发、生产、供应链与国际合作造成全方位冲击。中国企业面临的核心风险相互关联、容易形成"连锁反应"。

6.BIS Policy Statement on Controls that May Apply to Advanced Computing Integrated Circuits and Other Commodities Used to Train AI Models，https://www.bis.gov/media/documents/ai-policy-state-ment-training-ai-models-may-13-2025.
7.Know Your Cloud Customer: Commerce Department Proposes To Regulate Foreign Access to US IaaS Products，Skadden, Arps, Slate, Meagher & Flom LLP and Affiliates，https://www.skadden.com/insights/publications/2024/02/-know-your-iaas-customer.

（一）实体清单"关联风险"：从"单点打击"到"产业链牵连"

被列入实体清单的企业不仅自身面临"技术断供"，其上下游合作伙伴也可能因"关联关系"被BIS认定为"高风险企业"，导致"产业链整体瘫痪"。

风险表现：①上游供应商（如海外芯片设计公司）因担心"违规风险"，可能主动终止与清单企业的合作；②下游客户（如海外电子设备厂商）因"规避制裁"，选择取消与清单企业的订单；③物流、金融服务商因"合规压力"，可能拒绝为清单企业提供服务。

（二） 供应链"断裂风险"：从"设备断供"到"替代失败"

美国通过FDPR等规则，切断中国企业获取"关键设备、芯片、材料"的通道，而中国在部分领域（如EUV光刻机、高端AI芯片）的"国产替代"尚未成熟，可能导致企业面临"生产停滞"风险。

风险表现：①先进制程芯片断供：中高端AI芯片无法进口，企业大模型训练算力不足，研发周期延长；②制造设备断供：EUV光刻机、先进刻蚀设备无法进口，半导体制造企业无法推进先进制程研发；③材料断供：产品质量下降。

（三）投资交易"终止风险"：从"融资受阻"到"战略搁浅"

美国对外投资审查规则导致中国企业"海外融资渠道收窄"，同时，中国企业对海外企业的"技术并购"也可能因"美国干预"失败，影响其"技术获取"与"全球化布局"。

风险表现：①海外融资受阻：美国VC、PE因"审查限制"，不敢投资中国AI与半导体企业，导致企业"研发资金短缺"；②技术并购失败：中国企业对海外半导体、AI企业的并购，可能因美国"国家安全审查"被否决；③估值下降：因"投资限制"导致融资不畅，AI与半导体企业的海外估值下降。

（四）技术合作"中断风险"：从"协同创新"到"研发孤岛"

美国对高校合作、技术许可的限制，导致中国企业无法通过"国际合作"获取"前沿技术"与"创新理念"，技术迭代速度放缓，或形成两强并

立的国际产业格局。

风险表现：①高校合作终止：美国高校因"政策限制"，终止与中国企业的联合研发项目；②技术许可受阻：美国企业无法向中国企业转让"核心专利"；③标准参与受限：通过"国际标准组织"，如电气与电子工程师协会（Institute of Electrical and Electronics Engineers，IEEE）限制中国企业参与AI、半导体领域的"国际标准制定"，导致中国企业在"AI标准体系"中缺乏话语权等。

（五）合规"违规风险"：从"交易违规"到"天价罚款"

美国限制措施的"复杂性"与"模糊性"（如MEU规则中"潜在军事用户"的定义不明确），导致中国企业容易因"合规问题"触发"违规处罚"，且处罚力度大、影响范围广。

风险表现：①过失违规：企业因"尽职调查不到位"，误将"潜在军事用户"列为客户，导致出口违规；②规则理解偏差：企业因未准确理解FDPR规则，误将"含美国技术的海外产品"销售给实体清单客户，导致违规；③内部合规缺失：企业"出口合规体系"不完善，导致员工"擅自出口敏感技术"。

（六）品牌"声誉风险"：从"标签化污名"到"市场流失"

美国通过"实体清单""次级制裁"等措施，将中国AI与半导体企业污名化为"涉军""涉监控"企业，导致其"国际品牌声誉受损"，失去海外市场与合作伙伴。

风险表现：①海外客户流失：其他地区的客户因担心"合规风险"，拒绝与被列入实体清单的中国企业合作；②资本市场排斥：海外金融市场对被列入实体清单的中国企业要求特别的风险提示，导致资本市场运作受限；③公众形象受损：媒体通过"负面报道"，导致中国企业在海外的"公众形象受损"。

针对上述限制措施及中国企业面临的风险，笔者结合"法律合规+产业实践+国际资源"，提出"短期风险规避、中期能力建设、长期战略转型"的三级应对策略，帮助企业在"合规框架内"实现稳健发展。

（一）短期：建立"风险识别—阻断"机制，规避即时违规风险

短期核心目标是"避免触发美国制裁"，通过"客户审查、供应链筛查、交易合规"三大机制，阻断"违规风险点"。

（二）中期：构建"自主+多元"体系，降低对美依赖

中期核心目标是"减少对美国技术、设备、资本的依赖"，通过"国产替代、多元供应链、非美市场拓展"，提升企业"抗风险能力"。

1. 加速国产替代，提升"自主研发"能力

- 研发投入：加大对AI算法、半导体设备、材料的研发投入，重点突破"卡脖子"环节；
- 产学研合作：与中国高校、科研机构建立"联合实验室"，共同攻克核心技术；
- 专利布局：加强AI与半导体领域的"自主专利"布局，尤其是"核心算法专利""制造工艺专利"，减少对美国专利的依赖；同时，通过"专利交叉许可"与非美企业建立合作，获取"技术互补"优势。

2. 打造"多元供应链"，降低对美供应链依赖

- 供应商多元化：拓展非美供应商；
- 本地化生产：在亚洲、欧洲设立"组装厂"，利用"本地供应链"生产"不含美国技术的产品"，规避FDPR限制；
- 供应链联盟：联合上下游企业成立"供应链联盟"，共享供应商资源、联合采购降低成本。

3. 拓展"非美市场"，减少对美市场依赖

- 新兴市场开拓：重点开拓对AI、半导体产品的需求旺盛的"一带一路"沿线国家等新兴市场；

➢ 本地化合作：与非美企业成立"合资公司"，利用"本地品牌+本地渠道"拓展市场；

➢ 合规产品设计：针对不同市场设计"合规产品"，设计不含美国技术的性能产品，实现"市场差异化布局"。

（三）长期：推动"战略转型+国际协同"，构建"可持续发展"格局

长期核心目标是"跳出美国限制框架"，通过"技术标准输出、国际合规协同、政策支持对接"，构建"自主可控、开放合作"的发展格局。

➢ 参与AI"国际标准制定"，争夺"规则话语权"；

➢ 加强AI"国际合规协同"，应对美国"长臂管辖"；

➢ 对接"国家政策支持"，获取"资源保障"。

005 > 结语

美国对AI相关领域的管控已经从单一的出口限制，逐步扩展为覆盖出口、投资、合作与人才交流的全方位体系。这一体系的核心逻辑在于：AI不仅是技术竞争的前沿，更是国家安全的关键变量。对于中国公司而言，这意味着：

- 获取美国AI相关技术、芯片、算法或算力的难度将持续增加；
- 在境外投资、合作或学术交流中可能遭遇更多合规障碍；
- 美国政策的不确定性与动态调整需要持续跟踪。

然而，值得注意的是，美国的相关措施并非只针对中国，在其他国家和地区同样有应用案例。这表明，美国的管控思路更倾向于"风险导向"与"国家安全逻辑"，而非单一国家针对性。

在未来，AI技术的国际流动将越来越受到法律与政策的制约。企业在进行相关跨境活动时，应充分了解美国的法律框架与政策逻辑，避免因合规不足而带来不必要的法律和商业风险。

于治国
合伙人
合规与政府监管部
北京办公室
+86 10 5796 5075
yuzhiguo@zhonglun.com

作 者 / 王峰

算力之争：从成功帮助新加坡企业应对美国BIS调查H100转售和刑事起诉案件探讨中国企业如何思考和应对

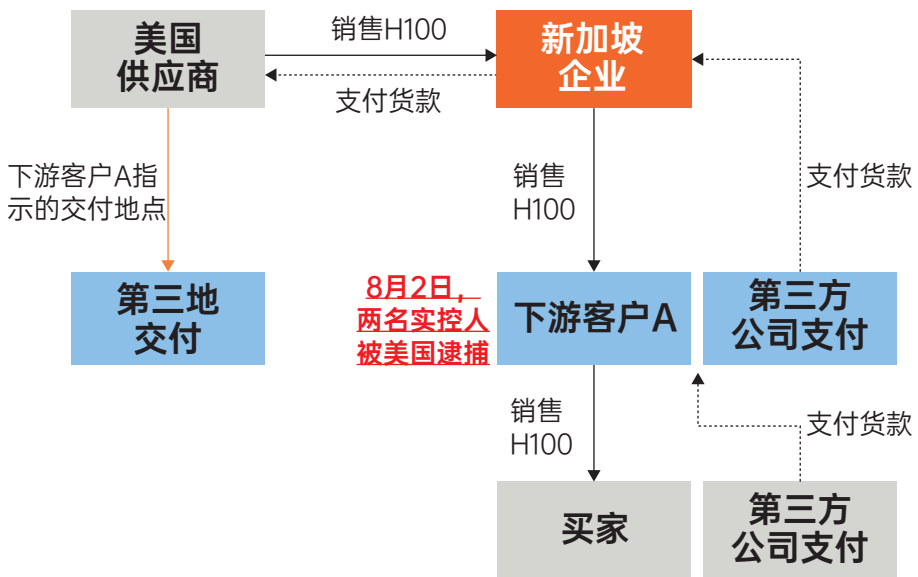
除了贸易战和关税战，中美在人工智能（AI）领域之争也越来越紧张。在人工智能领域的发展，简单而言主要体现在三个方面：算力、算法和数据。相比较算法和数据，美国在算力领域的整体能力的确更强，特别表现在算力硬件领先方面，如众所周知的英伟达高端算力卡的整体强悍运算能力。为了保持在算力领域的领先地位，美国通过制定法律法规，加强核查等手段，从限制算力相关技术和硬件输出至中国，到制造设备的禁售中国等方式，全方位阻断高端算力卡流向中国。从2018年至今，美国对中国算力领域逐步形成了"芯片-设备-技术-生态"的全面封锁。

中国很多企业都在试图获取更多更强的算力以帮助企业的模型升级迭代，在这个过程中我们注意到了很多问题，造成潜在违反美国出口管制和禁令的风险。所以，仅以2025年我们应对美国商务部工业与安全局（Bureau of Industry and Security，"BIS"）官员和新加坡海关官员调查的案例，帮助中国企业理解其中的一些问题（"坑"）和纠正错误理解，为企业健康、合规发展提供支持。

001> 案例背景简述和总结 – 企业需要关注案例中的相同点和重点，考虑自身业务场景中是否存在类似情况

如下图所示，新加坡企业（客户）从美国供应商采购H100服务器，转售下游客户A，服务器交付客户A指定的东南亚（如新加坡、马来西亚和泰国）收货企业，下游客户A自己或是通过第三方公司支付相关款项。

在今年8月2日，美国政府逮捕了下游客户A的两名实控人。从披露信息获悉，除了从新加坡企业采购H100，客户A还从其它企业采购H100，并转售至其他买家，货交东南亚（但按照美国政府披露内容，最终流向中国），货款由位于中国大陆和中国香港的第三方企业支付。



通过补全下游客户A的交易信息，进行对比可以发现，新加坡企业和下游客户A在业务场景中存在很多相同点：相同产品、相同转售商业行为、相同交付第三方、都有第三方香港公司支付货款情况。但最终结果却大相径庭，新加坡企业最终被认为符合美国出口管制规定和新加坡法律，而客户A的两名实控人被美国批捕。

这说明有很深层的情况需要中国企业重新思考和引起重视。如美国商务部BIS调查官通常思路是怎样的？如何提供证据证明企业的合规

性（即法律层面的抗辩）？从美国调查角度，哪些证据是必须提供的证据？哪些是最有力的证据？新加坡企业是AI领域相关的小规模企业，没有大企业的所谓完整的合规体系，如何向美国商务部BIS和新加坡海关解释其合规管控机制？公司的主要负责人还是中国国籍，且与中国国内公司有业务往来，美国官员会如何看待这种情况？且我们如何进行解释？

以此案件为基础，我们对美国针对中国的算力管控和调查做简略分析，提示整体特点，即：**法律规定的复杂性，违规风险的多样性，和业务方案和应对预案需要全面性**，希望帮助中国企业在处理相关业务中及早、及时和准确理解风险，做好事前、事中和事后的降风险安排，减少不必要的损失。

002 > 在AI算力领域，美国针对中国的限制内容非常复杂 – 影响到中国企业评估整体业务方案风险，切记不要自欺欺人

很多企业都在咨询关于AI算力相关的潜在违规风险问题，我们就先简述一下实务中的错误。举例来说，美国对中国禁售H100始于2022年10月，而H100产品上市也在同年10月。简单来说，H100基本就没有可能进入中国市场，进入中国市场的，很大可能都是违反美国禁令进口的。另外，产品技术指标也决定了产品的唯一性，那么通过技术指标也可以判断出产品类型，从而这也就满足了美国法律的"知道"中的"明知"。限于篇幅，我们不展开论述，但在此提示的是：法律规定需要结合到企业的订单、交付、ERP系统管理、议付资料等等内容，而非讲讲规定那么简单。仅从上述两个小问题，就可以看出规定本身与实务衔接的复杂性和潜在问题，以及规定本身与美国基本法律的衔接问题，更不用说美国还有至少10多个不同领域的政府部门牵头参与AI规定的制定。

特别强调的是：在今年2月份美国发布的**贸易优先政策**中，特别强调了现有出口管制中的漏洞问题，即通过战略对手和代理人的国家进行管制产品（如H100）转让。在**投资优先政策**中则强调了美国将对各种投资类型进行限制，以阻断对中国在**半导体、人工智能、量子、生**

物技术、高超音速、航空航天、先进制造、定向能源等领域以及中国国家**军民融合战略涉及**的其他领域的资金来源。

美国对AI领域的监管始于2018年底，在规划的14大类重点领域拟议适用出口管制中，人工智能就是其中重点之一，并从2019年开始，制定20多个限制规则，协调美国内部管理，针对外部事务进行管控。以下梳理，仅供参考。希望帮助企业充分理解规定的复杂性，而不是简单理解为"通过第三方"采购就不存在风险。

发布日期	发布机构	文件名称
2019年2月11日	美国总统行政办公室	《维护美国在人工智能时代的领导地位》
2019年5月15日	美国总统行政办公室	《关于确保信息通信技术与服务供应链安全的行政命令》
2020年3月12日	美国国会	《2020年国家人工智能倡议法案》
2020年5月1日	美国国会议员	《2020年数据问责和透明度法草案》
2020年5月19日	美国国会议员	《产生人工智能网络安全法草案》
2020年11月17日	美国总统行政办公室	《人工智能应用监管指南》
2021年1月19日	美国商务部工业与安全局	《确保信息和通信技术及服务供应链安全》暂行最终规则
2021年5月28日	美国国会议员	《2021年算法正义和在线平台透明度法草案》
2021年6月9日	美国总统行政办公室	《关于保护美国人的敏感数据不受外国敌手侵害的行政命令》
2022年3月2日	美国国会议员	《2022年算法问责草案》
2022年9月15日	美国总统行政办公室	《关于确保美国外国投资委员会有力地考虑不断变化的国家安全风险的行政命令》
2022年10月4日	美国总统行政办公室	《人工智能权利法案蓝图》
2022年10月7日	美国商务部工业与安全局	《对向中国出口的先进计算和半导体制造物项实施新的出口管制规则》

发布日期	发布机构	文件名称
2023年1月26日	美国国家标准与技术研究员	《人工智能风险管理框架》
2023年6月14日	美国国会	《保护美国人数据免受外国监视法案》
2023年10月17日	美国商务部工业与安全局	《先进计算芯片更新规则》《半导体制造物项出口管制临时最终规则》
2024年2月29日	美国商务部工业与安全局	《确保信息和通信技术与供应链的安全：联网汽车》拟议规则
2024年2月28日	美国总统行政办公室	《关于防止受关注国家获取美国人大量敏感数据和美国政府相关数据的行政命令》
2024年4月24日	美国国会	《保护美国人免受外国敌对势力控制应用程序法》
2024年4月24日	美国国会	《保护美国人数据免受外国敌对势力法》
2024年5月8日	美国国会议员	《加强海外关键出口国家框架草案》
2024年12月2日	美国商务部BIS	《增加外国生产直接产品规则和先进计算、半导体制造物项管控的改进》（IFR）
2025年1月13日	美国商务部BIS	《人工智能扩散框架》（IFR）
2025年1月16日	美国商务部BIS	《先进计算IC尽职调查措施、修订和澄清》（IFR）
2025年5月13日	美国商务部BIS	1)《关于可能适用于训练人工智能模型的先进计算集成电路及其他商品的管控政策声明》 2)《关于对中国先进计算集成电路适用通用禁令十的指南》 3)《防止先进计算集成电路转用的行业指南》

注：虽然个别规定已经失效，但理解内容可以帮助企业了解美国管控的思路和方法。

003 > 美国针对AI算力卡实施的调查 – 既有美国EAR法规支持，也有详细的程序和要求，切莫以为是走过场

仅梳理已经公开的信息，因为转售英伟达高端算力卡，在2025年已经采取行动的国家和相关信息如下：

日期	国家	部门	行为	刑事罪名	处罚	其他
2月26日	新加坡	警察部队和海关	突袭	虚假陈述实施欺诈	可判处最长20年监禁、罚款或两者兼施	还在调查22家企业和个人
3月13日	新加坡	新加坡检方和法院	开庭审理	指控新加坡公司欺诈性地掩盖目的地	同上	起诉3人，其中两名新加坡公民，一名中国公民
6月19日	马来西亚	**马来西亚投资、贸易和工业部(MITI)**	正在"核实"有关中国企业通过马来西亚绕过美国半导体限制的情况	暂无 根据马来西亚法律和事实情况决定，但强调"**任何个人或企业绕过出口管制都是非法的**"	暂无	/
8月2日	美国	美国联邦调查局(FBI)、美国商务部工业与安全局(Bureau of Industry and Security)	刑事调查中	涉嫌违反美国出口管制法规，采取绕道出口和规避监管的行为，非法将高级人工智能芯片——包括英伟达（Nvidia）H100加速器—出口至中国	针对个人最高20年监禁和100万美元罚款，可能还存在其他处罚，或监禁罚款两者兼施	/

除了上述严厉的刑事调查和起诉，在2024年，美国FBI、美国国土安全调查局（Homeland Security Investigations）和美国商务部BIS联合调查并于4月23日在芝加哥逮捕了两名中国公民。根据美国司法部公共事务办公室公布的一份起诉书披露，至少在2015年5月至2018年8月期间，此两名中国公民合谋逃避美国商务部对某中国企业实施的出口限制，方法是利用中介机构隐瞒实体清单企业参与交易的情况。起诉书指控的罪名是共谋非法出口美国技术，包括将美国加州公司生产的用于加工芯片的设备销售给了"实体清单"上的中国终端用户，违反了《国际紧急经济权力法》（IEEPA）和《出口管理条例》（EAR）。

相比事前的许可证申请制度，美国政府事后调查制度是企业需要更加重视的部分，因为从上述案例就可以发现，美国都是通过事后调查发现问题并追究企业责任的，而且往往都是"实锤"。我们在之前的文章《**直面美国BIS现场检查，思考企业合规实务**》着重介绍了相关程序和要求，这里不再作赘述。

简单总结就是美国在"转运后核查（Post-Shipment Verification）"中，流程规定非常严谨，要求参会人员非常全面，要求提供说明文件非常细致，而且是实地检查，企业不要有轻易过关的想法，因为企业提供任何形式的信息中，存在矛盾的地方都会被追问和反复确认。不专业和不专心应对都会造成企业巨大损失。这也就引出我们想分享的下一个话题：美国商务部BIS调查的思路。

004 > 美国政府调查的思路和方法 – 企业需要知己知彼，才能真正思考应对之策，"闭门造车"是自己坑自己

如果用一句话总结美国商务部BIS官员调查，那就是：非常认真、细致和全面，是综合能力的展示，体现了绝对的专业水准。

在调查过程中，BIS官员并不马上表态，但其从手中掌握的证据入手，希望被调查企业陈述自身的业务情况。以我们的案例为例，在确定企业的上游是美国知名服务器提供商后，BIS官员事先已经通过从供应商获得与新加坡企业交易的信息，再审核企业自己提供的信息，除了查看是否有隐瞒和出入外，还逐一比对企业手中的收货资料。

什么是逐一比对收货资料？收货资料有哪些？每一份文件上的每一个信息都代表什么，为什么要这样表述，和其他文件和信息的关联关系是什么等等，这些都是BIS官员关心的事项。我们之所以能够帮助到企业成功完成BIS和新加坡海关的调查，原因之一就是我们逐页、逐项梳理了全部资料，并总结成展示文件和提交文件，供BIS官员使用，而且充分阐述了业务中的瑕疵问题。

业务中有合规瑕疵，BIS官员会接受吗？在我们的案例中，企业的确有工作中的瑕疵，如接受了客户的要求，由第三方公司支付货款的行为。但这是国际贸易中的实务操作，第三方支付本身是没有问题的，而问题在于是否事先获悉并合理解释第三方支付的必要性和合理性。所以，商业合理性造成的瑕疵更容易被接受。这也是我作为高级经理在美国企业工作时，负责东北亚贸易合规工作而获得的实务经验。企业根本做不到100%的合规，但的确是已经在尽力的情况下，做不到的部分也就成为"不可能的任务"。当然，这需要具体问题具体分析，切不能自欺欺人式的自洽。

企业架构的问题需要重视。BIS官员还特别了解的新加坡企业的情况，包括负责人的人数，国籍等很多信息。在我们的案例中，企业的主要负责人现为中国籍，且还有国内还有研发团队，这些都是被调查内容。其目的就是从美国法律角度，判断新加坡企业的"法人独立性"和"总部"是否满足美国相关法律。而这些都是企业和外部机构忽视的地方。而且很多企业并不了解这些概念在实务中的表现形式。

企业合规体系是否是重要评估标准，这也是中国企业关心的问题，而且很多机构在对企业介绍时总结重点就是完善的合规体系。BIS官员和我们的思路是一致的，即实质大于形式，合规体系形式根本不重要。BIS官员并不纠结我们是否能够拿出完善的承诺、政策、流程等等文件，而是关注于实际执行层面的情况。这也是非常好理解的，因为美国商务部BIS公开处罚中国山东某企业，就是因为美国政府认为一些企业"说漂亮话，做违规事"，"表面一套，背后一套"。所以，我们也一直提醒企业要注意实际执行层面的细节。而这些细节又从企业提供的各种信息中得到相互印证，进而说明问题。正所谓"细节是魔鬼"，可以披露出本质，让BIS官员抓住企业是否违规的实质性内容。

在与美国商务部BIS和新加坡海关官员沟通中，还有很多思路上的特点，但总结如上，就是非常专业和细致。企业面对美国商务部BIS官员的调查，表面是运输后的事后调查，实际上是从企业性质、交易模式、交付和支付、上下游管理、法律规则的知晓等各个方面进行全方位的调查。

005 > 中国企业应该考虑的事项

事项1：从微观角度，处理好细节，如此案中对"转售"与"欺诈"的理解，还有业务中涉及的其他重要概念

"转售"是企业商业活动中的一种常态行为，都是新加坡企业，都是转售，但认定结果却大相径庭，我们的客户新加坡企业合规，下游客户A违规并被刑事逮捕。这说明在转售过程中，对于正常转售与商业欺诈的边界需要有清晰的理解。

举一反三，那么企业无论是出海，还是投资、并购等都需要对业务场景中的重要概念做好充分理解，如原产地、税则归类、工艺改变、加工与制造的区别、明知和应知、责任穿透等等，不仅仅是中国法律规则下的理解，还有如美国等国外法律规则的理解。对企业来说，正确理解是做好应对方案的前提。

事项2：从整体规划和应对角度，因为美国调查非常专业，如果涉及此类型业务，怎样做好业务方案和应对预案

从上述复盘"美国政府调查思路"中，我们可以看到美国商务部BIS官员的专业性和调查的全面性和细致性，既有企业架构，管理方面的信息，又有细节内容的逐一比对核实。所以，企业需要在各个方面做好事前的评估，避免出现矛盾而引发违规问题。

做好业务方案和应对预案的前提是要充分了解美国出口管制规定和调查思路，这样才能不违规。这就提示了另外一个问题，企业如何了解美国规定。

在客户A刑事调查中，由于H100是美国货物导致的美国法律在域外管辖的问题，其"欺诈"的界定也会考虑美国出口管制的影响。这就是实

务中需要考虑的法律问题。

事项3：正确理解美国规定的建议 – 多学习、多研究、多比较，并结合业务场景

通常，企业的做法就是自己研究和咨询外部机构，如律师事务所，咨询机构等。而自己研究也会通常会咨询外部机构。这就出现了一个问题，外部机构是否专业，是否能够为企业提供非常全面和细致的信息。那么企业如何判断外部机构是否专业？结合我们的经验，建议企业在考虑外部机构选择时，思考以下内容：

1)从事过10年以上企业内部业务管理工作。真正从事业务工作比外部机构参与某些项目要务实和细致很多。参与处理过一些项目的经验远远不及多年从事业务工作的经验丰富和务实。10年以上企业实务工作经验也会帮助企业从战略和战术角度更全面的思考和提供意见；

2)直接掌握和理解美国法律。如果在美国企业工作、被美国律师培训、与美国律所对抗（非合作）等，这些都可以直接获得美国法律的细节和美国企业的实务操作细节，从而避免因为理解偏差而造成违规风险；

3)行业内比较知名的企业的经验，如一些头部企业，有被美国调查的企业，他们的经验会比外部第三方只讲规则要务实和有用。

事项4：VIE架构，开曼架构，壳公司都不是保险方案 – 美国追责穿透至自然人

从我们的案例和获悉的下游情况，我们可以引申思考，就是从中国企业角度，设立海外公司，利用VIE架构等作为商业上的框架安排是否可以帮助企业控制风险。企业架构是美国调查的重要内容之一，因为美国法律（包括美国EAR）追责的总体要求就是要穿透到自然人，追究违反美国规定的背后负责人。而美国法律下的"法人独立性""总部"的规则，就是避免企业钻空子。所以中国企业无论是VIE架构设计、或是在东南亚投资、设厂、设立公司等等以阻断风险的方法，需要谨慎处理，而不是简单设立一个公司就可以解决追责问题。

事项5：业务方案和应对方案要有前瞻性和预判性

我们曾经在2018年底就提示美国针对中国企业"绕道"避税的措施，也提示过美加墨自由贸易协定并不能达到完全帮助中国企业通过墨西哥从而减轻税负，在Trump1.0时代，就帮助了东莞企业不转移生产线而到达调整原产地从而减少税负。

在最近的项目中，企业还咨询我们的方案是否可以持续2-3年，这也是我们建议企业思考的问题之一，即无论怎样的方案，都需要有生命力，都需要不仅考虑当下的限制，业务场景，还需要考虑将来业务发展的情况以及准确预判法律规则的变动。如下述事项6，我们如果在境外开展工作，就要预判第三国是否遵守美国法律的问题，并提前做好准备。

事项6：第三国是否遵守美国法律的问题

在新加坡正式逮捕转售H100的相关人员前，根据CNA的报道，**新加坡贸易与工业部第二部长**Tan See Leng在2月18日的国会上表示，新加坡不纵容企业故意利用与该国的关系来规避或违反其他国家的出口管制。此外，**马来西亚**当局也开始调查这一案件，评估是否违反了马来西亚的相关法律。

无独有偶，在2024年12月2日，**马来西亚贸易部副部长**在一个论坛上已经表示："在过去一年左右的时间里.....我一直在警告许多中国企业，如果它们只是想通过马来西亚**更换产品标签，以避开美国关税，就不要投资马来西亚**。"

以这种级别官员做上述表述，传递信息已经非常清楚了。在美国法下，对于被认定为设立"壳公司""中间商"而从事违反美国法律的行为，都是美国政府重点和严厉打击的"绕道"行为。而从上述事件可以看出，针对中国可能违反美国相关法律的"绕道"行为，东南亚国家已经开始考虑进行必要行动了。那么中国企业在投资和开展业务前就要做好评估和方案，投资后要做好应对核查的预案。

总结梳理一下，中国算力需求是刚需，美国方面是非常清楚的。在预期5-10年中，中国肯定会迎头赶上补齐算力硬件方面的差距，但在这个期间，美国方面肯定会祭出杀手锏，尽最大力量阻止中国算力发展。就像这些美国在东亚和东南亚针对H100的调查，可以想象在其他地区

也有类似的调查进行中。而第三国被迫采取措施也是大概率事件。

就像此次美国商务部BIS适用50%规则（现暂停实施），表面上强调只是按照股权判断，但实际上，50%实控权才是美国各个领域法律的关键，不能排除美国商务部可能是留了后手，会根据中国控股企业股权变化而准备后续精准打击。所以中国企业需要跳出自己思维的圈子，更认真，更仔细，更全面的思考如何降风险，这不仅仅在AI领域，也是在贸易战、关税战、科技战，先进制造战等等可能的挑战中需要重新思考的。

王峰
合规与政府监管部
北京办公室
+86 10 5796 5001
wangfeng9@zhonglun.com

chapter

03

知识产权与数据治理 | 人工智能创新

intellectual property and data governance : ai innovation

作 者 / 张鹏 牟雨菲

统筹设计：人工智能创新的多维度知识产权布局体系

未来已来，人工智能技术浪潮席卷全球，传统的知识产权制度正面临前所未有的挑战与机遇。为适应这一变革，2025年4月30日，国家知识产权局发布《专利审查指南修改草案（征求意见稿）》（以下简称"修改草案"）和修改说明，突破性地将相关章节规定，从原"包含算法特征或商业规则和方法特征的发明专利"，调整为"涉及人工智能等的发明专利"。这一变革显著拓展了人工智能技术的专利保护客体的范围，表达了有关部门对于人工智能技术专利保护的积极态度。从创新主体知识产权保护体系规划的角度，应积极利用客体保护的变革，综合运用专利、商业秘密和著作权等多种法律工具，形成互补协同的保护策略。本文将从修改草案出发，探讨人工智能技术的知识产权多维保护要点，以及如何借鉴药品专利的多维度保护策略及专利生命周期管理经验，最后，在全球视野下规划人工智能创新保护策略的简要思路，为企业在人工智能时代的创新竞争提供参考。

001>人工智能专利保护客体边界以及修改草案的潜在影响

2025年4月30日，国家知识产权局发布《专利审查指南修改草案（征求意见稿）》和修改说明，将原第二部分第九章"包含算法特征或商业规则和方法特征的发明专利申请审查相关规定"的标题及范围修改为"涉及人工智能等的发明专利申请审查相关规定"，并增加了针对人工智能发明专利申请依照专利法第五条的审查基准和示例、典型案例所说明的创造性审查规范以及满足说明书充分公开的撰写要求和示例。

（一）修改草案的基本背景及修改内容

通常而言，专利制度秉承"公开换垄断"的基本逻辑，在授权确权规则中具有三重"门槛"，只有通过这三重门槛的专利申请才能够得以授权成为授权专利。这三重"门槛"是，第一，是否属于专利法保护的客体，是否适宜运用专利制度加以保护（事实上只有少数创新成果属于运用专利制度加以保护的创新成果）；第二，"公开换取垄断"中的公开是否充分（说明书是否充分公开）、垄断权保护范围是否清楚（专利权保护范

围是否请给出）以及所公开的内容与所垄断的范围是否匹配（权利要求是否能够得到说明书的支持）；第三，专利权利要求是否具备创造性。具体而言，修改草案针对人工智能专利审查的修改内容主要围绕贯彻落实国家政策驱动、引导和规范创新实践、统一和明确审查标准[1]三大修改必要性展开，针对上述三重"门槛"对人工智能技术专利申请授权确权规则加以完善，旨在建立更清晰的法律边界、提升专利申请质量，并统一审查标准。

1.增加有关专利法第五条有关违反法律情形下的审查规定，引导智能向善

修改草案明确规定，涉及计算机程序的发明专利申请必须接受合法性、道德性审查。修改草案"2.涉及计算机程序的发明专利申请的审查基准"中删除审查针对权利要求记载的解决方案的规定，明确涉及计算机程序的发明专利申请需要针对A5进行审查。对于人工智能、大数据专利申请，如果方案本身目的违法或一旦实施必然违法，则属于专利法第五条违反法律的情形，不授予专利权。例如，个人信息采集必须符合《个人信息保护法》等法律规定，必须以维护公共安全所需或取得个人单独同意为前提。

2.增加说明书应充分公开的具体撰写要求，引导申请质量提升

针对人工智能领域普遍存在的"黑匣子"问题，本次修改对说明书的撰写提出了更为具体的要求。在"6.3.1说明书的撰写"部分的第二段中，在明确说明书应当写明算法特征如何与技术特征共同作用并且产生有益效果的后面，以"又如"的形式，在原有"例如"后，增加有关人工智能模型训练或构建、人工智能应用类相关申请在说明书撰写上的具体规定。修改草案明确区分了两种常见情形：对于涉及人工智能模型构建或训练的专利申请，要求详细说明模型的训练方法、数据来源及处理过程；对于人工智能模型的应用类专利申请，则需要清晰阐述算法如何与具体技术特征相结合，并产生何种技术效果。

1.详见：国家知识产权局《指南第二部分第九章第6节内容修改说明》。

3.完善创造性评判标准，通过示例统一审查边界

为解决人工智能领域创造性判断的实践难题，本次修改特别增加了正反两个审查示例，厘清了创造性评判的边界。在"6.2审查示例"的"（4）在进行创造性审查时，应当考虑与技术特征在功能上彼此相互支持、存在相互作用关系的算法特征或商业规则和方法特征对技术方案作出的贡献"中，增加正反两个示例，旨在说明对于大数据、人工智能领域专利申请，在识别对象不同时，创造性如何把握。

在修改草案的两项审查示例中，正向案例"一种建立废钢等级划分神经网络模型方法"表明，当处理对象的特殊性（如废钢的复杂形态）导致必须对模型结构、训练方法等进行针对性调整，且这种调整与技术特征形成协同作用时，算法特征应当被纳入创造性考量。相反，"一种船只数量确定方法"的反向案例则说明，如果仅仅是将现有算法应用于新的数据对象，而未对算法本身带来实质性改进，则不能认定为具有创造性。

（二）修改草案的影响及意义

人工智能发明创造是以算法为基础、在"大数据"与"大计算"的共同驱动下融入多技术领域、不同功能维度的多项单一技术方案所形成的综合性技术束[2]，因此，其算法的权利要求可能被认定为智力活动的规则和方法[3]、数学方法[4]、或计算机程序等情形，从而面临各国传统专利法律制度的排除规则是否适用、如何适用，相关人工智能专利适格性是否应予排除等争议问题。

2019年12月31日，我国发布《关于修改<专利审查指南>的公告》[5]，明确人工智能发明创造的可专利性及创造性审查标准。在第二部分第九章专门新增的第6节"包含算法特征或商业规则和方法特征的发明专利申请审查相关规定"中，针对包含人工智能算法的发明专利申请的专利适格性审查，即按照以下步骤和规则进行：（1）根据《专利法》第

2.刘鑫、覃楚翔："人工智能时代的专利法：问题、挑战与应对"【J】，载于《电子知识产权》2021年第1期。
3.《专利法》第 25 条。
4.ibid. [2]
5.国家知识产权局公告第343号。

二十五条第一款第（二）项审查权利要求是否属于“智力活动的规则和方法”。（2）根据《专利法》第二条第二款审查权利要求是否属于“技术方案”。针对包含人工智能算法的发明专利申请，首先从反向视角审查其是否属于专利适格性排除主题即智力活动的规则和方法，若该申请没有落在专利适格性排除主题范围内，则进一步从正向视角审查其是否构成专利法意义上的技术方案。只有同时通过正反向视角的两步审查，人工智能算法发明专利申请才具有专利适格性。并且，在审查包含人工智能算法的发明专利申请时，坚持整体审查原则，即不应当简单割裂技术特征与算法特征，而应将权利要求记载的所有内容作为一个整体[6]，对其中涉及的技术手段、解决的技术问题和获得的技术效果进行分析。

总体来看，本次修改草案在人工智能发明专利的创造性审查标准基础上，对人工智能相关专利申请的审查规则作出了进一步细化。一方面，明确了涉及个人信息采集与算法伦理等场景的审查底线，强调技术方案须符合法律与社会公德。另一方面，通过新增创造性审查示例，为专利撰写提供了实践指引，要求突出算法特征与技术特征的协同作用，说明书必须打破"黑匣子"模式，详细公开模型结构、训练步骤及数据关联，以确保技术方案的可实现性。从说明书公开的角度，修改草案规定在人工智能相关申请的说明书中，必须清晰记载模型的必要模块、层级或连接关系，训练所需的具体步骤与参数，以及模型与场景的结合方式、输入与输出数据之间的关联等。

002>综合运用知识产权体系构建人工智能保护组合策略

在人工智能创新领域技术迭代迅速，市场竞争激烈，单一的知识产权保护模式往往难以覆盖人工智能创新的各个方面，需要在现行法律制度框架下开展知识产权保护。人工智能创新企业应灵活运用专利、商业秘密和著作权等多种工具，形成协同效应，以实现对AI技术成果的多维度的有效保护。

6.国家知识产权局10720号复审请求审查决定。

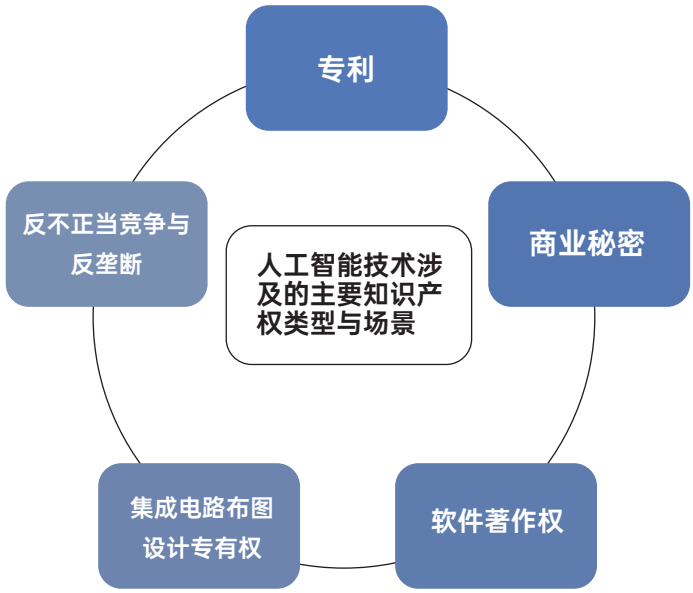


图1 人工智能技术涉及的主要知识产权类型与场景

（一）人工智能创新的专利保护

人工智能相关专利申请的解决方案通常涉及人工智能算法或模型，以及人工智能算法或模型的功能或领域应用，主要以发明专利的形式进行申请。人工智能技术分为基础层、感知层、认知层、应用层四个层次，需要实现立体式的专利布局。基础层是实现大计算驱动和大数据保障的基础算法，感知层主要体现为语音技术、图像技术、视频技术、AR/VR增强现实基础等感知性技术，认知层主要体现为人工智能涉及的自然语言处理、知识图谱、用户画像等以机器学习为核心的认知性技术，应用层主要是无人驾驶、智能制造等应用场景。

专利权的核心性质为"禁用权"，专利权人通过公开相关发明的技术方案获取专利权，从而能够就未获得合法授权的专利实施者主张侵权责任，以司法或行政手段推动侵权方达成许可或制止其实施行为。专利维权方式方面，主要有民事诉讼、行政保护、刑事制裁三大类型，各路径遵循独立的程序法则又存在有机衔接，共同构建起专利权的执法保障网络，此外涉及标准必要专利的，还有可能涉及标准必要专利反垄断投诉。根据《专利法》第六十五条规定，"未经专利权人许可，实施其专

利，即侵犯其专利权，引起纠纷的，由当事人协商解决；不愿协商或者协商不成的，专利权人或者利害关系人可以向人民法院起诉，也可以请求管理专利工作的部门处理”。

对于人工智能发明创造而言，由于专利权的本质在于以公开换取保护，技术方案的披露尺度本质上是法律保护形态的战略抉择。在满足专利所要求的充分公开和支持的前提下，概括出人工智能的算法或模型及其改进、优化、应用，例如，模型结构、模型压缩、模型训练等，并尽量少披露具体技术方案的落地核心技术细节，如参数配置、算法实现细节、训练数据等有关内容。此类内容在专利中进行披露和限定，反而会导致权利要求的保护范围缩小，使竞争对手更易规避，对于该价值密度高、容易形成竞争壁垒的高质量技术构成可能采用技术秘密保护更为合适。

（二）人工智能创新的商业秘密保护

商业秘密，是指不为公众所知悉，具有商业价值，并经权利人采取相应保密措施的技术信息、经营信息等商业信息。我国《民法典》明确将商业秘密列为知识产权的客体。其中，技术信息是指利用科学技术知识、信息和经验获得的技术方案，包括但不限于设计、程序、公式、产品配方、制作工艺、制作方法、研发记录、实验数据、技术诀窍、技术图纸、编程规范、计算机软件源代码和有关文档等信息。

对于人工智能创新的商业秘密保护而言，应基于专利权与商业秘密的不同权利特质进行的融合，对不同技术方案进行结构，一方面专利通过技术方案的充分公开换取法定期限内的市场垄断，另一方面，通过商业秘密权对无法通过反向工程破解的核心算法技术细节、参数组合及数据进行保护。

具体而言，应对策略往往采用专利与商业秘密相结合的模式，即在专利申请文件中，仅公开实现技术效果的必要信息，将最优参数范围、特定数据预处理技巧等关键细节作为商业秘密保留。这种模式既利用专利划定权利边界，又通过商业秘密守护技术内核，形成刚柔并济的保护格局。计算机软件著作权则作为补充，保护技术方案表达形式的独创性，开源策略则代表另一种选择，通过特定许可证（如GPL、

Apache）主动放弃部分排他权，以技术公开换取生态共建与标准主导权，组合策略需在核心闭源与外围开源的平衡中实现，将基础框架开源以吸引开发者，同时通过专利与商业秘密锁闭核心数据训练系统与迭代算法。

技术解构与组合保护绝非静态的法律安排，而是伴随技术与法律环境持续演进的动态过程。在技术维度，人工智能的快速迭代特性要求保护策略具备前瞻适应性。研发初期作为商业秘密保护的参数调优方法，可能因技术变迁而失去保密价值，此时需评估转化为专利的可能性。而原本处于专利保护核心的硬件设计，可能被新一代技术取代，则可考虑进行开源以维持行业影响力。在市场维度，全球化布局需应对法域差异的复杂挑战，一方面考虑在专利保护强度高的地区，可侧重专利布局，而在反向工程盛行的市场，则需强化商业秘密保护体系。竞争态势的演变同样影响法律策略，当遭遇专利狙击时，可启动专利无效宣告程序或以自有专利组合反制，积极寻求律师协助；发现商业秘密侵权迹象时，需迅速采取证据保全措施；面对行业标准形成窗口期，可通过有选择地开源抢占生态位，涉及标准必要专利时积极进行FRAND承诺。这种动态调整要求创新主体建立知识产权与研发、市场部门的跨职能协同机制，使法律策略深度嵌入技术创新全生命周期。

（三）人工智能创新的著作权保护

人工智能创新的实现方案在著作权保护方面均存在一定空间，亦存在不足之处。就人工智能创新的实现方案而言，由于思想表达二分法下仅仅保护作品的表达，使得软件著作权对人工智能基础算法的保护非常有限，著作权仅能保护代码的文字表达，却无法覆盖算法背后的设计思想与架构。

通常，人工智能算法的软件化实现需经历三个关键环节：首先是需求分析与架构设计，系统分析员通过深入调研，明确系统功能模块划分和界面设计，并完成系统的整体架构设计，包括处理流程、模块功能分配、数据结构等核心要素；其次是详细设计与编码，开发者在架构基础上进行细化设计，明确算法实现、类结构关系等具体方案，并据此编写实现各模块功能的程序代码；最后是测试与发布，通过系统测试确保软

件质量，并完成相关文档编制。从这三个环节的产出成果来看，第一阶段形成的是软件架构设计，第二阶段产出的是程序代码，第三阶段则形成测试文档。显然，软件著作权只能保护第二环节中的代码表达，却无法保护体现核心创造力的软件架构。而随着现代软件开发工具日益智能化，软件架构的设计价值正愈发凸显，这使得著作权保护的局限性更加突出。

因此，在上述专利与商业秘密结构性知识产权保护策略的基础上，人工智能的底层算法软件自生成之日起自动受到著作权保护，但在实际主张权利的过程中，一般作为商业秘密的组合权利基础进行维权保护。

003>借鉴药品专利策略实现人工智能保护周期管理

在保护范围及保护周期的布局策略方面，人工智能技术的立体式专利布局可以借鉴药品的"化合物专利——组合物专利——制备方法专利——变换性专利——用途专利"的立体式专利布局模式[7]。如图2上半部分所示，药品创新链一般涵盖靶标确立、生物学模型建立、先导化合物研发与优化、临床前及临床研究、新药申请与批准以及药品转用等环节。

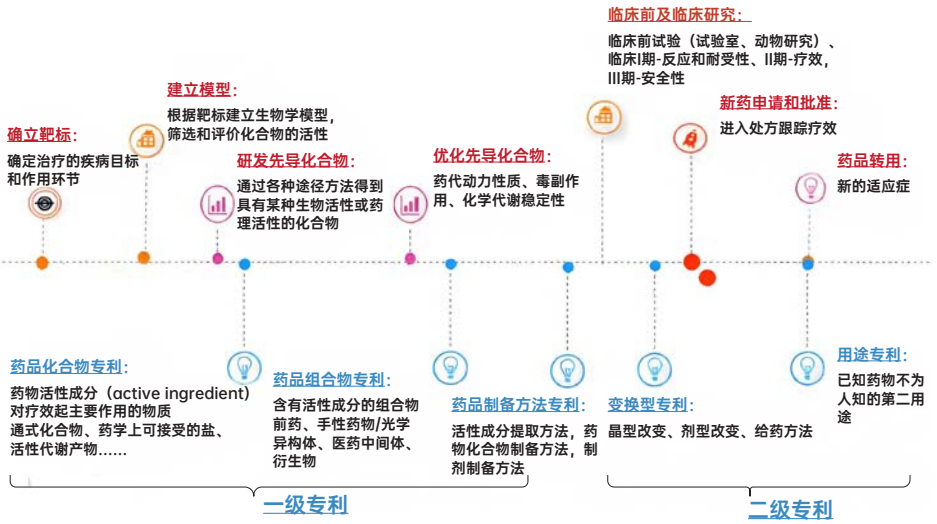


图2 可以作为参照的药品专利立体布局模式

7.邱福恩："人工智能算法创新可专利性问题探讨"【J】，载于《电子科学技术》2020年第1期。

其中，确立靶标、建立模型、研发与优化先导化合物被视为研究阶段；临床前及临床研究、新药申请与批准、药品转用则属于开发阶段。整个流程统称为"药品创新链"。"确立靶标"是起点，确定疾病治疗目标与作用环节；随后通过构建生物学模型（如药代动力学模型）筛选和评估化合物活性。接着是研发先导化合物——即具备特定生物活性及化学结构的化合物，其来源包括天然活性物质、现有药物的副作用改进及合成中间体筛选等。由于先导化合物常存在活性不足、药代性质差、毒性大或稳定性低等问题，通常需进一步优化。完成研究阶段后，进入开发阶段：开展临床前研究（实验室或动物试验，评估活性与安全性）、临床研究、新药申请与批准，以及药品转用。

与创新链对应，药品专利布局链如下：在靶标确立与模型建立后，通过专利检索确定先导化合物，并申请先导化合物专利，保护通式化合物、药用盐及活性代谢产物，构成药品最基础的专利。由于相同结构的药物可能因结晶条件不同形成多晶型，影响药效与质量，因此在优化过程中常形成组合物专利，保护由多种成分按比例构成的具有特定性质和用途的混合物。例如，辉瑞公司在降脂药阿托伐他汀的专利布局中，先后申请了保护通式化合物的US4681893和保护其钙盐的组合物专利US5273995。随后可申请制备方法专利，涵盖提取、分离、纯化及制备工艺。化合物、组合物与制备方法专利形成于研究阶段，创新程度高，被视为一级专利。在临床及审批阶段，则可能产生如晶型等变换型专利；在药品转用阶段则形成用途专利，如新医药用途或新适应症。变换型与用途专利属于二级专利，虽基于一级专利再创新，但创新程度未必低。例如，西地那非原用于冠心病疗效不佳，后发现可治疗勃起功能障碍（CN94192386X）与肺动脉高压（EP1097711），即为典型的新用途专利。这样的权利要求即属于《专利审查指南》第二部分第十章第4.5节规定的"用途权利要求"，亦即将基于发现产品新的性能，并利用此性能而作出的发明。[8]

8.详细讨论参见张鹏："抗击疫情药物的用途专利申请前景与合规使用探析——以瑞德西韦专利布局分析为视角"【J】，载于《中国发明与专利》2020年第2期。

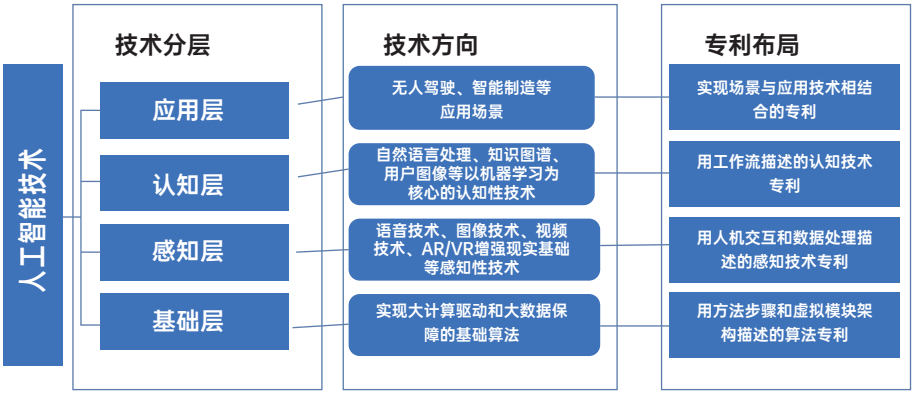


图3 人工智能技术立体专利布局模式

将药品的阶段性专利保护措施应用于人工智能创新的延长保护，应建构"基础算法专利——感知技术专利——认知技术专利——应用场景专利"共同构成的人工智能技术专利布局。如前所述，人工智能技术通常包括基础层、感知层、认知层、应用层四个层次，基础层是实现大计算驱动和大数据保障的基础算法，感知层主要体现为语音技术、图像技术、视频技术、AR/VR增强现实基础等感知性技术，认知层主要体现为人工智能涉及的自然语言处理、知识图谱、用户画像等以机器学习为核心的认知性技术，应用层主要是无人驾驶、智能制造等应用场景。

这其中，基础算法构成的基础层是整个人工智能技术的基础和核心，然而传统的专利法律制度认为算法属于智力活动的规则和方法，从而人工智能技术理应被排除在专利法的保护范围之外[9]。算法的专利保护，无非是在现有专利授权确权标准之下，根据促进人工智能发展的需求，划分出具有可专利性、可以授予专利权的"技术方案"和不具有可专利性、不能授予专利权的"智力活动的规则"[10]。因此，需要通过撰写加工等方式，促进基础算法专利与应用场景或者感知技术、认知技术的结合，形成对基础算法的专利布局。特别需要注意的是，在寻

9.张洋："论人工智能发明可专利性的法律标准"【J】，载于《法商研究》2020年第6期。
10.孔祥俊："人工智能知识产权保护的若干问题"【J】，载于《上海法学研究（集刊）（2019年第13卷 总第13卷）——上海市法学会互联网司法研究小组论文集》。

求基础算法的专利保护过程中，需要与应用场景相结合时，必须认真分析基础算法除了当前应用场景之外的其他可能应用场景，对"场景替换式"的侵权行为进行有效规制。

感知层主要体现为语音技术、图像技术、视频技术、AR/VR增强现实基础等感知性技术，可以采用人机交互和数据处理的方式加以描述，从数据流流向的角度总结处理流程形成方法权利要求，从模块架构出发形成装置权利要求。认知层主要体现为人工智能涉及的自然语言处理、知识图谱、用户画像等以机器学习为核心的认知性技术，可以采用工作流的方式从工作流流向的角度总结处理流程形成方法权利要求，从模块架构出发形成与方法权利要求对应的装置权利要求。应用层主要是无人驾驶、智能制造等应用场景，类似于上述药品专利的用途权利要求，将特定基础算法、特定感知层和认知层的工作流、数据流处理方案与应用场景进行结合，对特定应用场景下的使用进行保护。

004 > 人工智能知识产权的国际布局与保护现状

在国际布局层面，人工智能创新的知识产权保护布局主要在于海外落地专利的全球布局策划，而各国对于人工智能创新，尤其是人工智能算法发明创造的可专利性存在不同程度的审查标准，对人工智能创新的在各国专利局申请增加难度和阻碍，企业需要根据不同法域对于人工智能作为专利客体保护的审查标准及程度，就不同法域的专利申请进行调整，同时最大程度避免减损相关发明创造在不同法域的保护范围。

英国法院在*Emotional Perception AI Ltd v Comptroller-General*一案[11]中认定，当人工智能算法发明属于计算机程序，若其解决了计算机外部技术问题，提供了一种外部技术效果，则其并非对计算机程序本身的权利要求，亦应当属于专利保护客体的范围。本次修订实质性地拓

11. [2023] EWHC 2948 (Ch).

宽了人工智能发明保护客体的范围。

欧洲专利局（EPO）在实践中，多将专利新颖性以及创造性作为审查重点。对于人工智能的"创造性步骤"，审查焦点为，其核心技术特征（多数为计算机程序或基础算法）必须贡献于该发明的技术特性，从而使得涉案技术具有技术效果。对于这一要求，欧洲专利局审查委认为，涉案技术应通过其应用于技术领域来达到技术目的或通过适用于特定技术应用来实现技术效果[12]。因此，一般获权人工智能技术主要为特定领域的技术应用，通用型人工智能以及纯粹在基础算法层面进行突破的相关技术难以通过EPO专利审查[13]。但EPO对通用性AI或基础算法实现技术贡献的人工智能专利适格性持积极态度，通过充分具体说明训练数据和所处理的技术任务，人工智能神经网络也被认定为对自动化人工任务或解决技术问题提供了特定的技术应用。[14]

美国针对人工智能技术可专利性判定采取拟制现有技术排除测试法，将涉及抽象概念的部分拟制为对专利新颖性和创造性不具有任何贡献的现有技术，在新颖性和创造性判断中加以排除。根据美国法司法实践，美国最高法院明确了不授予专利权的客体包括自然规律、物理现象和抽象概念[15]。美国最高法院2014年*Alice*案形成了"拟制现有技术排除测试法"的基本逻辑，将上述自然规律、物理现象、抽象概念拟制为对专利新颖性和创造性不具有任何贡献的现有技术，在新颖性和创造性判断中加以排除，要求权利要求的其他部分具备新颖性和创造性。[16]对于人工智能技术发明专利而言，尤其需要判断是否属于"抽象概念"，亦即如何区分受到专利法保护的包含算法特征或商业规则和方法特征的发明专利和属于抽象概念的不属于专利法保护的创新创造。这是由于算法本身更类似于数理逻辑，而与解决技术问题的技术手段存在一定差异。

日本方面，对于人工智能的专利审查主要分为两步，第一步判断方案是否利用自然规律。在第一步无法确定的情况下，进行第二步，

12.T 2330/13.
13.《欧洲专利审查指南》Part G Chapter II 第 3.3.1 节。
14.ibid, p.20.
15.*Diamond v. Chakrabarty*, 444 U.S. 303, 206 U.S.P.Q. 193.
16.狄晓斐："人工智能算法可专利性探析——从知识生产角度区分抽象概念与具体应用"【J】，载于《知识产权》2020年第6期。

首先审查是否实现对设备的控制或基于物理/化学/生物/电学特性处理信息，之后要求审查员根据"软件是否通过硬件资源具体实现"来确认是否达成"利用自然规律的技术思想"，核心是特定信息处理器或其操作方法需基于预期用途实现软硬件协同。人工智能基础算法与应用场景相结合的方案通常认为属于可以受到专利法保护的"技术方案"。日本特许厅在2018年3月出台《面向人工智能相关技术的审查指南实例》，结合具体案例给出了人工智能基础算法与应用场景相结合的方案的可专利性审查标准。其中，对于人工智能算法与应用场景相结合的发明创造，明确属于专利权的保护客体。《面向人工智能相关技术的审查指南实例》认为，如人工智能发明技术方案利用硬件资源实现了软件的信息处理，属于可以受到专利法保护的"技术方案"。针对生成式AI的兴起，日本特许厅发布了含10个案例的新指南，其中两个生成式AI案例明确了创造性的判断标准：单纯应用AI替代人工任务的系统化设计（如客服应答生成器）缺乏创造性；相反，权利要求2因包含了"在预定字符限制内生成与问题高度相关且适宜作为参考的附加文本"这一特定技术手段，而被认定具备了创造性。

韩国方面，KIPO在审查专利适格性时采用四步法，其中关于抽象算法规则的审查重点一般在第三步：软件的信息处理是否通过硬件具体实现。KIPO发布的《AI相关发明审查案例》如果用于实现AI发明的技术方案中包含区别技术特征（例如，数据预处理、训练模型具有区别技术特征）、使用AI输出的结果数据中包含区别技术特征、AI的应用产业领域完全不同等，并且区别技术特征带来了更好的技术效果，则发明容易被认为具备创造性。

世界知识产权五大局（即中国国家知识产权局CNIPA，美国专利商标局USPTO，欧洲专利局EPO，日本特许厅JPO，韩国特许厅KIPO）均在放松人工智能专利审查规则，尤其是专利客体适格性的方面。欧专局在五大局中要求显得最为宽松，美国依赖判例，日韩仍旧相对保守。以先前经验来看，五大局的审查规则势必会引领全球人工智能专利审查标准的主要趋势。

在人工智能技术的全球专利布局中，应基于企业所处的行业进行具体分析。首先应当通过PCT途径进行国际专利申请，在此基础上，

根据技术特性和市场战略，科学选择进入国家阶段的路径。布局范围不仅需要覆盖传统的主要技术市场，还应重点关注新兴技术区域的发展态势。对于关键技术，要确保在核心市场获得充分保护；对于辅助技术，则可根据市场前景进行选择性布局。在技术方向选择上，专利布局需要首先重点关注模型训练与优化相关的核心技术，包括神经网络架构、机器学习算法等基础技术，其次，应当聚焦企业所处应用场景的行业在不同地域的人工智能领域的关键技术节点，例如数据隐私保护、身份认证等安全技术的专利布局。此外，还需要特别关注输入输出数据处理、模型与场景结合等应用层技术的专利保护，通过对关键技术节点的系统布局，形成对整体技术方案的立体保护。

张鹏
高级顾问
知识产权部
北京办公室
+86 10 5957 2068
zhangpeng@zhonglun.com

作 者 / 赵刚 蒲柯洁

浅谈人工智能模型的法律保护——全国首例未经许可使用他人模型结构与参数案件述评

2025年3月31日，北京知识产权法院对原告北京抖音科技有限公司诉被告亿睿科信息技术（北京）有限公司侵害著作权及不正当竞争纠纷一案作出二审判决，维持一审北京市朝阳区人民法院认定被告未经许可使用原告"变身漫画特效"模型构成不正当竞争，但不构成著作权侵权的判决。[1]本案亦是全国首例针对未经许可使用他人模型结构与参数行为作出侵权认定的生效判决。

001>案例简介

（一）事实概要

原告为抖音APP的运营者，开发"变身漫画特效"并上线于抖音APP，该特效可以将用户拍摄的真人照片、视频转换为漫画风格，被用户广泛使用。被告为B612咔叽APP的运营者，在抖音APP上线"变身漫画特效"之后，B612咔叽APP上线"少女漫画特效"。两款APP内置争议特效功能相似，同时处理相同照片能够得到相似的结果；作为对比，案外人开发的其他人像动漫化平台处理相同照片得到的结果与两款APP争议特效处理结果存在显著差异。

原告主张其开发之变身漫画成像是美术作品、视听作品，同时主张变身漫画特效对应之模型结构和参数系其核心竞争优势，受到反不正当竞争法保护，认为被告行为构成侵权，要求法院判令被告停止侵权、消除影响、赔偿损失500万元及合理开支24万余元。

原告开发特效对应模型的过程分为：①风格设定②风格化量产③模型训练④用户生成之四个阶段。①风格设定阶段，确定变身漫画成像标准为"像"（相似）和"美"。②风格化量产阶段，聘请手绘师对照Faceu软件[2]公开的5万余张真人照片以前一阶段设定的风格为标准绘制漫画。③模型训练阶段，在CycleGAN模型[3]的基础上调整结构与参数并使用前

1.北京知识产权法院（2023）京73民终3802号民事判决书。可参见知产宝：《典型案例 | 全国首例！北京知识产权法院：未经许可直接使用他人训练模型的结构与参数构成不正当竞争行为》，https://mp.weixin.qq.com/s/0ryGs8-97p0sObhjwpl-jg.
2.由原告关联公司开发运营的自拍APP。
3.一款开源图像到图像转换模型，例如可以实现将图片中的马转换为斑马、风景图季节由夏天转换为冬天等。模型介绍详见https://github.com/junyanz/CycleGAN。该模型适用较为宽松的许可证，仅要求保留版权声明、许可证和免责声明即可再行分发该模型，无论是否进行了修改。判决中显示的本案最终选用的pix2pix模型亦采用相同条件。

一阶段产生的真人照片和手绘师绘制的漫画进行训练，过程中反复调整模型结构、参数和输入真人照片和漫画的成对数据（如由手绘师手动修复模型无法调整的漫画瑕疵等）。④用户生成阶段，将训练好的模型供抖音APP调用，用户通过抖音APP使用"变身漫画特效"。该模型如输入内容相同则输出结果大致相同。

专家意见指出，案涉原告模型针对特定日本漫画风格迁移任务作了特定设计，与其他GAN模型在结构等方面存在明显不同，而原被告模型除部分细节存在对输出结果存在微小影响的差异外基本一致，原被告同时设计出相似模型的可能性较小。被告未提交充分的证明其独立研发和原被告模型存在实质性差异的证据。

（二）法院认为

关于著作权侵权，本案一审法院结合《著作权法》及其实施条例中"创作"的概念指出，著作权法意义下的创作行为不是单纯积累素材、数据、创造生成工具的行为，也不能是按照特定规则机械完成工作、缺乏创作空间的行为。前述第②阶段之风格化量产阶段属于积累数据的工作，第③阶段之模型训练阶段属于创造生成工具的行为；而第④阶段中，用户使用该特效生成的内容与真人存在唯一或有限的局限性，无法体现人的思想、情感与个性。因此，原告在四个阶段中均没有进行著作权法意义上的创作行为，变身漫画成像本身并非著作权法保护的客体。二审中由于原被告双方未对著作权问题提出上诉，二审法院未对著作权问题进行评价。

关于不正当竞争。二审法院在归纳适用《反不正当竞争法》第2条认定不正当竞争的要件的基础上进行了分析。其一，竞争关系上，原被告均通过APP为用户提供视频和照片的拍摄服务，具有直接竞争关系；其二，被诉行为不属于知识产权专门法保护及反不正当竞争法第二章规定的具体不正当竞争行为的情形；其三，被告使用了原告的模型结构及参数，并且该行为违反了人工智能模型领域公认的，不得未经许可直接使用他人通过数据训练改进而来的模型结构和参数的商业道德；其四，原告为案涉模型投入大量经营资源，模型为其取得了创新优势、经营收益和市场利益，原告享有受反不正当竞争法保护的合法权益；其五，原

被告案涉产品效果相似，存在较强替代作用，被告对原告竞争利益造成了实质损害；其六，被诉行为扭曲了模型正常供求机制，如不规制则将助长"搭便车"行为，无法恢复扭曲的供求机制和创新机制，使市场激励机制失灵，扰乱人工智能模型领域的竞争秩序。二审法院基本确认了一审法院关于被告行为构成不正当竞争的认定，并据此维持一审法院判决。

002 > 人工智能模型的反不正当竞争法保护路径

本案中，法院适用《反不正当竞争法》第2条对原告主张的模型结构和参数进行了保护。在司法实务更多关注人工智能模型训练数据和生成内容本身侵权的当下，对于模型结构与参数本身的评价将具有更显著的意义。二审法院详细地列出了适用第2条进行保护的要件，并基于本案事实进行了具体分析，为人工智能模型的保护角度和方式开创了示范性先例，笔者认为亮点如下：

首先，法院肯定了人工智能模型开发者所应享有的反不正当竞争法上的合法权益的边界，即竞争利益不能通过给予宽泛的竞争优势或交易机会来推断，而是要具体结合人工智能模型开发者的实际开发投入、效果功能进行综合评价。这就要求人工智能模型开发者应当在维权过程中详尽对开发过程、开发投入以及基于模型结构和参数的应用产生的产品效果进行充分举证，而不能仅仅从产品本身推导出竞争性利益。

其次，法院以判决的方式明确了人工智能模型领域的商业道德标准，即"从事人工智能模型研发经营的企业不得未经许可直接使用他人通过数据训练改进而来的模型结构和参数"，本质上对人工智能领域"不劳而获"行为进行了否定性的商业道德评价。

第三，在行为事实是否成立的判断上，接触+实质性相似的基本法则同样适用于人工智能模型结构和参数比对领域，同时对于自主研发的举证责任同样归属于被告承担。

第四，实质性替代作用依旧是合法权益是否被损害的基本判断准则。只不过笔者这里需要强调的是，基于法院裁判的逻辑，实质性替代作用的判断不应拘泥于模型结构和参数本身，而是应当延伸至基于该模型结构、参数的特效产品。因为在判断用户群体、目标市场、产品提供

途径等方面，只有商业化的产品才能发挥出价值衡量的尺度，而模型结构、参数的保密特性导致其价值衡量往往需要通过产品本身得出。

第五，人工智能模型市场竞争秩序的判断需要结合行为对市场运行的准入机制、供求机制、价格机制、信息机制、信用机制、创新机制等的影响进行综合判断，“搭便车”“不劳而获”型的侵权行为，往往对价格机制和创新机制的损害是最大且最为直接的。

那么这里可能会有一个疑问，本案为什么不考虑商业秘密的《反不正当竞争法》保护角度呢？目前对于人工智能模型的保护，似乎以商业秘密角度保护的呼声是最高的。笔者认为，本案通过商业秘密保护可能存在一个无法绕过的障碍，即在无证据证明被告系通过非法或员工跳槽等方式获取模型结构或参数数据的情况下，唯一相对合理的解释是“技术破解”，这也恰恰印证了法院的调查事实，即根据专家意见“技术人员可以通过技术手段提取模型本身并进行解密，从而获得模型的结构和参数”。而技术性破解往往会落入“反向工程”的范畴[4]，这恰恰是商业秘密侵权案件中法定的可抗辩胜诉角度之一，原告没有选择商业秘密的保护路径至少对于本案来讲是十分合理且审慎的。

003 > 人工智能模型的著作权保护路径

本案原告在著作权保护上主张了模型输出结果的美术和视听作品的保护角度。对此，一审法院完整地评价了模型开发应用全阶段的各行为均非著作权法意义上的创作行为，并未予以支持。如在②风格化量产阶段，虽然输入的数据系用户拍摄的照片，可能具有一定独创性，且手绘师依据真人照片绘制的漫画单独来看同样可能具有一定独创性，进而受到保护，然而相关工作整体对于模型本身而言属于“积累数据的工作”。再如④用户生成阶段，用户在具体使用案涉特效时，本身是拍摄照片或视频的过程，该过程产生的照片或视频可能具有独创性，然而用户拍摄真人照片或视频也与成像效果存在唯一或有限的对应性，成像本身无法

4.虽然学界可能对此有不同的声音，但笔者仅从审慎角度予以推断原告维权角度选择的可能性供读者参考。

体现人的思想、情感与个性。笔者赞同法院关于著作权问题的论述，但如果抛开原告主张的权利作品类型，而仅从模型角度思考，模型本身是否可以类比计算机软件作品予以保护呢？这确实是一个值得深入思考的问题。

计算机软件作品本身在著作权法各例示作品中的地位便并不寻常——计算机软件本身就是极为特殊的一类作品。计算机软件的作用不是展示文学、艺术和科学美感，传递思想感情或特定信息，而是指挥计算机完成任务，对于其保护方式早期亦存在较大争议。[5]美国Boudin法官曾指出"将版权法适用于计算机程序就像是在拼接一个各部分之间无法适应的七巧板。"[6]假设将真人照片转换为漫画的工具不是人工智能模型而是传统的计算机软件（例如Photoshop等传统计算机软件也可以通过风格化滤镜等传统图像操作实现真人漫画化），则开发该类计算机软件的过程也属于创造生成工具的行为，此时除非出现算法与计算机语言表达的混同，则恐难以否定该类计算机软件受到著作权法保护。从用户角度看，在"黑箱"的处理过程中，对于类似的输入，模型和传统计算机软件都能给出一定的输出，如此讨论模型与传统计算机软件的区别便存在价值。

如前文所述，计算机软件是指挥计算机完成任务的工具。在生成式人工智能迅猛发展之前，传统计算机软件通常由人类手动编写。亦即，人类根据计算机程序语言和编码规范将算法转写为源代码形式，对于固定的输入，传统计算机软件可以给出固定或符合预期的输出；但对于人工智能模型，则更多是事先以大量的输入和输出使得人工智能模型拟合出逻辑的过程。

视语言性质，传统计算机软件中的源代码可能编译为二进制目标程序，对于目标程序，虽然人类不可直接轻易识读但其与源代码相对应，法律对此给予等同保护。《计算机软件保护条例》第3条第1项对此有明确规定。模型和目标代码同样作为人类难以轻易识读的二进制文件形

5.王迁：《知识产权法教程》（第七版），中国人民大学出版社2021年版，第135页。
6."Applying copyright law to computer programs is like assembling a jigsaw puzzle whose pieces do not quite fit." See 49 F.3d 807, 34 U.S.P.Q. 2d (BNA) 1014, 1995 WL 94669, 1995 U.S. App. LEXIS 4618.

式，对于模型结构部分，模型开发的结构设计阶段是高度要求设计者创造性和智力的阶段，暂且不提模型的最终使用需要传统计算机软件的配合，模型这一二进制文件同样需要通过传统源代码等表达形式定义结构，虽然在可逆性与确定性上存在差异且模型包含参数部分，这一过程存在类比为源代码编译为目标代码的过程可能性。司法当前热议的提示词与生成内容关系的案例或亦可作一定类比：如承认详尽的、具有独创性的提示词生成的内容可以受到保护，则定义结构的代码到模型文件的过程或许也存在一定相似之处。

而对于参数部分，笔者试以简单的数学函数打一个极为粗略且不尽精确的比方：传统计算机软件中，人类程序员首先选定了一种结构形式（如$y=f(x)=ax+b$），由人类程序员通过计算机程序语言告知计算机参数a与b究竟为何，计算机软件即可根据每次不同的x的输入给出y的值的输出；然而在人工智能模型层面，对于人类首先选定的结构（此处同样使用$y=f(x)=ax+b$粗略举例，实际远比该示例复杂），人类以大量成对的x和y进行训练以便其拟合出合适的a和b的值，最终得到二进制形式的模型文件。基于著作权法对计算机软件作品保护的代码表达确定性视角来看，训练的结果和编写的结果相比，模型参数确实存在著作权法层面保护的弱势。

诚然，根据著作权法原理，著作权法不保护算法本身而保护程序本身的表达（无论是源程序还是目标程序）。即使是传统计算机软件，其最具价值的设计思想和实用性功能也未得到著作权法层面的保护，对于模型而言本身亦是如此，但亦以计算机软件作品代码保护形式类推对其结构进行一定程度的保护，以防止他人接触后使用，可能存在一定的可行性和探讨空间。

（一）开源模型问题

传统开源许可证本是针对传统计算机软件设计的，虽然也存在将其用于计算机软件作品之外的其他作品的情况，但将其用在模型之上也存在一定贴合性。[7]如前文所述，本案原告选用的开源模型适用的许可证本身条件非常宽松。一审法院以被告未提交证据证明涉案开源模型属于强著作权许可类型，因此对被告关于GAN模型的答辩意见未予采纳。[8]

传统软件即使涉及输入和输出，其输出权利归属亦通常不存疑问地与输入一致。在生成式人工智能迅猛发展且AIGC权利归属规则尚不完全明朗的当下，CreativeML Open RAIL-M等许可证专门针对模型的许可作出了安排，并设置了关于模型输出内容权属的特别约定。在传统计算机软件视角下，开源许可证系对著作权许可问题作出安排疑问不大[9]，但由本案延伸讨论，在本案适用《反不正当竞争法》第2条时二审法院归纳的各要件下，其他违反开源许可证使用他人开源模型的行为是否可能被评价为违反人工智能模型领域的其他商业道德？对权利人维权会导致什么样的后果？也是非常值得思考的问题。

（二）模型训练数据合法性问题

全球视域内已有多起案件涉及数据权利人起诉模型开发运营企业未经授权使用其数据。我国被称为“全球首例生成式AI服务侵犯著作权的生效判决”的AI形象侵权一案中，AIGC平台应原告方“生成A”的简单提示词要求即生成出A的形象，平台最终被法院判决侵权。[10]该案涉及训练数据本身侵权问题，权利人基于训练数据本身主张权利。然而，假设有充分的证据证明模型基于存在权利瑕疵的数据训练，其模型的结构与参数是否还能得到保护呢？

7.许可人完全可以选择GFDL及Creative Commons系列许可证用以开放许可文字作品等传统作品类型，但将用于计算机软件的开源许可证用于文字作品等其他传统作品类型的情况，可参见如https://www.gnu.org/licenses/gpl-faq.html#GPLOtherThanSoftware，并非完全不可行。

8.虽然我国司法实务针对开源许可证的讨论通常均针对GPL等具有传染性、条件较为严格的许可证展开，但对于仅具有署名等类似宽松要求的许可证在违反时同样可能存在达到解除条件成就或合同自始不生效的效果的空间。

9.例如广州知识产权法院（2019）粤73知民初207号民事判决书认为GPL是附解除条件的著作权许可格式合同。

10.广州互联网法院（2024）粤0192民初1xx号民事判决书。

本案中被告辩称原告适用非法获取的数据进行训练获得的竞争性利益不应受到保护。一审法院认为该问题不影响本案评价，二审法院指出被告未提交证据证明，人脸照片数据与人工智能模型的结构和参数之选择和使用直接相关，进而均未支持被告这一抗辩。

在传统著作权法领域，即使是基于他人作品进行未经授权的改编，法律依然承认未经授权改编后的作品权利存在，未经许可他人不得使用。虽然模型开发训练者完全可以购买或通过其他渠道获取，并使用权利完全无瑕疵的类似质量的成对数据得到相似的模型结构与参数，且模型结构与参数本身不会包含训练中使用的成对数据，但训练数据与模型本身结构重要性不相上下的语境下，使用有权利瑕疵的数据可能使得模型训练者得到节省开发训练资源的"搭便车"的效果。瑕疵训练数据是否将对其竞争性利益产生足以否定该利益存在的影响，可能仍需个案具体衡量。或许可以类比的是，在传统著作权法语境下，改编作品确实能够形成新的作品权利内容，但基于公平诚信角度考虑，改编作品的使用同时也不得侵害原作品权利人的合法权益。基于这个角度，瑕疵训练数据训练所形成的人工智能模型或参数本身或许仍旧能够获得反不正当竞争法项下的权益内容之司法认可，但瑕疵可能将对其维权或多或少产生行权障碍，比如无法获得侵权禁令甚至赔偿等，对此人工智能模型的开发企业应当值得关注。

赵刚
合伙人
知识产权部
北京办公室
+86 10 5087 2893
zhaogang@zhonglun.com

作 者 / 顾萍 王成荫 伍波

创新与著作权的平衡：Thomson Reuters v. ROSS Intelligence 案

对AI训练数据的规制

人工智能与知识产权法的交叉领域正迅速发展，并在全球范围内变得至关重要。近期，美国特拉华地区法院审理的"汤森路透企业中心有限公司等诉罗斯智能公司"（*Thomson Reuters Enterprise Centre GmbH et al. v. ROSS Intelligence Inc.*，以下简称'*Thomson Reuters v. Ross Intelligence*'）一案，是美国在利用受著作权保护的材料进行AI训练方面具有里程碑意义的判决。考虑到中国作为全球AI发展领域的领先者，此案的判决结果及其法律逻辑无疑具有深远的全球影响。本文旨在介绍此案，解释其关键方面，并分析其对中国AI产品研发活动中，尤其是在训练数据来源方面的法律风险和启示[1]。

001>案件背景

*Thomson Reuters v. Ross Intelligence*案件源于法律数据库内容被用于人工智能训练的纠纷。原告汤森路透公司（Thomson Reuters）运营大型法律数据库Westlaw，收录了几乎美国所有法院判决，并由编辑团队为每个判决要点撰写主题提要（headnote），配以相应的键号体系（Key Number System）分类。主题提要是对判决法律要点的简明陈述，键号则是按法律议题给每个提要编号，方便律师查找处理相同法律问题的其他案例[2]。被告Ross Intelligence公司是一家法律人工智能初创企业，试图开发一款AI法律检索工具，让用户用自然语言提问，快速提供相关的法律案例、法规和分析，从而简化传统的法律研究流程[3]。为实现这一功能，Ross需要大量训练数据来进行训练。

Ross公司最初曾请求许可使用Westlaw数据库的数据来训练其AI，但遭汤森路透拒绝。随后，Ross与第三方咨询公司LegalEase Solutions合作，使用LegalEase提供的大量法律问题摘要和相关案例，称为"Bulk Memos"，供AI训练使用。对于Bulk Memos的部分内容，LegalEase提供的Bulk Memos在内容和结构上与原Westlaw数据库的

1.本文分析引用了 *Thomson Reuters Enterprise Centre GmbH v. Ross Intelligence Inc.*, No. 1:20-CV-613-SB (D. Del. Feb. 2025) 判决书及相关评述。
2.Westlaw, Wikipedia, https://en.wikipedia.org/wiki/Westlaw (last visited Mar. 19, 2025).
3.参考*Features, Ross Intelligence*, https://www.rossintelligence.com/features (last visited Mar. 19, 2025).

内容高度相似。汤森路透发现后起诉Ross侵犯其主题提要和键号体系的著作权[4]。

本案的裁判要点在于：未经授权复制他人数据库中受著作权保护的内容用于AI训练，是否构成侵权；美国著作权法下"合理使用（fair use）"的抗辩是否适用于这种情况？

002>合理使用原则

本案涉及的核心法律是美国著作权法（在本文中，"copyright law"译为"著作权法"）及其合理使用原则（17U.S.C.§107）。合理使用允许在特定情况下未经授权使用受保护作品，例如为了评论、教育等目的[5]。案件由第三巡回上诉法院法官Bibas（临时指派担任特拉华联邦地区法官）审理。在各方提起的简易判决动议中，法官先是倾向于将合理使用交由陪审团认定，但后来重新审视证据后改变立场，改为由法官直接裁定合理使用问题[6]。2025年2月，法院作出部分简易判决，认定Ross未经许可使用Westlaw编辑内容训练AI不属于合理使用，构成对汤森路透著作权的侵犯。

判断是否构成合理使用，法院要综合考虑四要素[7]：

- 用途和性质(包括使用是否具有商业性及是否具有变革性[8])；
- 受保护作品的性质；
- 使用量和实质性；
- 对原作品市场或价值的影响。

Ross公司提出合理使用抗辩，认为其对数据的使用具有变革性目的（训练AI属创新）且最终产品并未包含原文，从而应受§107保护。不仅如此，Ross还质疑这些主题提要是否有足够独创性受著作权保护，并提出"无意侵权""著作权滥用"等抗辩。

4.需要指出，美国判决书原文属于公共领域，但Westlaw添加的编辑性内容（如主题提要）具有独创性，因此受《美国著作权法》保护；Ross并未直接向用户展示这些提要文本，而是将其作为训练AI的数据集。
5.Fair Use, Digital Media Law Project, http://www.dmlp.org/legal-guide/fair-use (last visited Mar. 19, 2025).
6.Stuart D. Levi et al., Court Reverses Itself in AI Training Data Case, Skadden (Feb. 2025), https://www.skadden.com/insights/publications/2025/02/court-reverses-itself-in-ai-training-data-case.
7.17 U.S.C.§107.
8.Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994).

在综合四要素后，法院进行了利益权衡。Bibas法官明确指出：第一和第四要素支持原告，第二和第三要素支持被告。但鉴于第二要素权重相对较轻，而第四要素至关重要，整体平衡倾向于认定不构成合理使用。法官在简易判决中裁定Ross未经许可复制汤森路透编辑内容用于AI训练已构成直接侵权，Ross关于合理使用的抗辩以及著作权滥用、并入理论（merger）等抗辩均不成立。需要说明的是，法院裁定的部分简易判决意味着在法律责任上Ross已被判侵权成立，但仍有个别事实问题留待审理，例如部分提要著作权是否已过期无效等。

003>关于"变革性使用（transformative use）"的认定

法院在分析合理使用的第一个要素——使用的目的和性质时，着重讨论了"变革性"（transformative）使用概念。按照美国最高法院最近在*Andy Warhol Foundation v. Goldsmith*案中的解释，如果被诉二次使用与原作品的目的"相同或高度近似"，且属于商业性质，那么除非有其他正当理由，第一区分因素将倾向于不构成合理使用。换言之，只有当新使用赋予原作品新的目的或不同性质（further purpose or different character）时，才具有变革性。在本案中，Bibas法官认为Ross对汤森路透内容的使用缺乏变革性：Ross利用主题提要训练AI，其最终目标是提供一个功能近似的法律检索工具，与Westlaw平台本身的用途并无二致。正如判决中指出的，Ross拿这些提要来更容易地开发一个竞争性的法律检索工具。Ross的AI是将用户的法律问题与已有判决相匹配，这与Westlaw利用提要和键号检索相关案例的过程极为相似。因此，从直观上看，Ross的使用目的和汤森路透原作品的用途高度重合——都是为了让法律从业者更便捷地检索判例。法官据此认定Ross的使用并非变革性的，因为它的目的和性质并没有进一步的目的或与汤森路透的作品有不同的性质。这一判断直接指向第一区分因素不利于合理使用成立。

值得注意的是，Ross辩称其复制行为发生在产品开发的中间阶段，最终用户并不会看到Westlaw的提要原文。这类似于软件领域常见的"中间复制"问题，即为了实现互操作性而临时复制受保护代码的情形。确有先例支持中间复制在特定条件下属于合理使用，例如第九巡回

上诉法院在*Sega v. Accolade*案和*Sony v. Connectix*案中分别认定：为开发兼容游戏或设计新平台，对计算机程序代码进行中间复制是变革性的使用，因其目的是实现新品功能，与原用途不同。此外，2021年美国最高法院在某版权纠纷案中也认为，被告复制原告API代码用于手机平台具有合理使用性质，因为API接口本质上功能性强，复制对实现新的兼容用途是必要的。然而，Bibas法官明确区分了上述案例与Ross案：首先，上述案例涉及的是计算机代码的复制，而本案复制的是文字性作品（编辑性法律内容）。法律在判断合理使用时，对待功能性程序代码与文学作品是不同的——程序往往以功能为导向，包含非表达的功能要素，而文字作品主要承载表达性创作。正如法院引用上述2021年美国最高法院某版权纠纷案所言：计算机程序不同于图书、电影等文学作品，其几乎总是服务于功能目的，因此针对程序的合理使用考虑不一定适用于文字作品。其次，即便在软件案中，法院也是基于"复制的必要性"来认定中间复制合理：例如2021年美国最高法院某版权纠纷案中被告复制API是不同程序相互对接所必需的；Sony和Sega案中被告若不复制代码就无法获取非保护要素或实现产品兼容。而Ross案并不存在必须复制汤森路透提要才能训练法律AI的技术障碍——Ross完全可以自行阅读公开的判决并编写自己的摘要来训练模型，并非别无他法。正如判决中指出的，汤森路透创建的任何内容，并非Ross不能靠自己创造或通过合法手段获得的。综合以上，法院认为Ross将原告编辑内容用于相同目的的商业行为并不具备变革性，其试图以"中间技术步骤"来为复制开脱的理由也不成立。Bibas法官在判决书中坦言，此前他曾据Sony/Sega案例倾向于将第一区分因素交由陪审团判断，但经过更深入分析后认识到这些软件中间复制案例与本案差异明显，本案应回归最高法院在*Warhol*案中强调的"目的对比"框架。他总结道，Ross获取提要只是为了更方便地开发一个直接与Westlaw竞争的法律检索工具；所以Ross的使用不具变革性。由此，第一个因素（包括商业性和变革性）整体倾向于原告汤森路透，不支持合理使用。

004>其他合理使用要素

在认定第一区分因素不利于被告后，法院继续分析了其余三个要素。第二要素（作品性质）方面，Westlaw的主题提要和键号属于编辑性整理作品，虽具有一定独创性，但主要围绕公开的法律事实和理念，创作性相对有限。Bibas法官指出，这类法律注释作品的创造力显著低于纯文学艺术作品，因此在合理使用判断中，不应像高度原创作品那样给予强保护。他认定第二区分因素偏向Ross一方（有利于合理使用抗辩），但也提醒该因素在以往案例中"很少起决定性作用"（在"电子图书馆"一案中，法院亦曾评价第二要素通常权重不高）。

第三要素（使用的数量和实质性）上，Ross通过LegalEase实际复制了数千条主题提要用于训练，这无疑是大规模全文复制。表面看，大量复制倾向不利于合理使用；但法院更关注向公众提供了多少原作内容。Ross的最终产品（AI检索结果）给用户看的并不是Westlaw的提要，而是相关联的司法意见全文，且这些法院判决本身不受著作权保护。也就是说，Ross复制汤森路透内容仅作为内部训练之用，并未向公众输出这些受保护表达。根据"电子图书馆"案的原则，应考量实际向公众提供了多少原作内容，以及这种提供是否可能成为原作的替代品。在本案中，用户无法通过Ross的产品获取汤森路透的编辑性表述，因此第三要素总体上对Ross有利。法官也驳回了Ross关于"只用了Westlaw提要库一小部分"的辩解，指出即使复制占比不高，若涉及作品精华部分仍可能过度（正如*Harper & Row v. Nation*案中非法引用福特总统回忆录的300字已被视为取走作品"核心"）。不过在Ross案，法官最终认为由于未向公众提供提要文本，第三要素可以判定偏向Ross。

*Campbell v. Acuff-Rose*等先例确立了第四要素——对原作市场或潜在市场的影响——是最重要的考虑因素。法院需评估被告的使用是否替代了原作在现有市场的需求，或妨碍了权利人开发衍生市场的可能性。Bibas法官指出，汤森路透原作品的现有市场显然是法律检索平台服务，而潜在衍生市场则包括将编辑内容授权用于AI训练的数据市场。他回顾自己在先前意见中曾有顾虑，认为Ross的产品也许服务于一个不同用途的新市场，不一定构成Westlaw的替代品，并且当时不确定汤

森路透是否有意涉足AI训练数据市场。然而，重新审视事实后，法院认为这些顾虑并不存在：即便按对被告最有利的事实看，Ross开发该AI工具的意图就是打造Westlaw的市场替代品，争夺相同的客户群。Ross自己也承认其产品直接与Westlaw竞争。至于汤森路透尚未将提要内容商业化为AI训练数据包并不重要——关键在于此类潜在市场本属于权利人合法预期范围，侵权人不应通过不授权复制来抢占先机。法官强调，哪怕原告尚未进入该衍生市场，只要这种市场可能存在且会受影响，第四要素也应认定不利于被告；举证说明这些市场不存在或不受影响的责任在于被告，但Ross未能提供足够证据。此外，Ross主张其产品有助于公众更便捷获取法律信息，应被视为对公共利益有利。然而法院指出，美国司法判决文本本身是自由公开的，公众有权免费查阅法律原文，但公众无权要求获取汤森路透对法律的解析。正如判决书中所言："**著作权法鼓励人们开发有益社会的事物，例如优质的法律检索工具。这些开发者有权因此获得报酬**"。也就是说，法律需要在公共利益与激励创作之间取得平衡。允许Ross不付费直接拿走汤森路透辛勤编辑的成果来牟利，会损及此平衡。不仅公共利益主张不足以豁免，Ross复制行为实际上损害了汤森路透应有的收益。法院进一步区分了2021年美国最高法院某版权纠纷案案情：在该案中，被复制的API之所以重要，是因为大量用户习惯于该接口，属于软件行业特有情形；而Ross完全可以独立创造类似的法律要点摘要来训练AI，并不存在必须使用汤森路透提要的情形，因此对原告权益的侵蚀缺乏正当性。综上，第四要素明显倾向于原告汤森路透，认为Ross的使用对原作品现有及潜在市场都有负面影响。

总体而言，该判决从司法层面对AI训练数据的著作权使用边界进行了划定，即：若AI训练使用他人受著作权保护的内容，且用途与原作品市场相竞争，则难以被认定为合理使用。

005>与中国法律的对比分析

与美国开放性的合理使用原则不同，中国著作权法采用封闭式的法定许可与合理使用列表。中国《著作权法》第24条列举了若干种无需许可的使用他人作品情形（合理使用），例如个人学习研究、课堂教学、新闻报道等，但并未包含"大数据/AI训练"的情形。这意味着，在中国，像Ross这样为商业目的大量复制他人作品用于AI训练，一般不在法定的合理使用范围内。我国业界也注意到这一立法缺口。有学者建议利用著作权法第24条中的兜底款，通过修改实施条例增设"数据训练"的合理使用例外，并辅以"三步检验法"限制，以平衡技术创新和著作权保护。不过截至目前，这仍只是理论建议，相关法律尚无明文规定[9]。

虽然中国尚未出现像"*Thomson Reuters v. Ross Intelligence*"案这样直接涉及AI训练数据著作权的判例，但中国法院在AI生成内容的著作权问题上已经做出了一些重要的裁决。

例如，北京互联网法院在2023年11月的一起案件中裁定，AI生成的图像如果体现了人类在提示和参数选择方面的智力投入，则可以享有著作权。法院认为，通过设计角色呈现方式、选择和安排提示以及设置相关参数等方式，原告投入了一定的智力劳动，使得生成的图像并非仅仅是机器的机械产物，而是体现了原告的个性化表达，因此构成受著作权法保护的作品。这一判决与美国版权局的立场形成对比，后者通常要求作品必须由人类创作才能获得著作权[10]。然而，广州互联网法院在2024年2月的一起案件中裁定一家AI公司因提供用户生成奥特曼图像的服务而侵犯了著作权[11]。法院认为，AI生成的奥特曼图像与原著作权作品高度相似，侵犯了原告的著作权。该案还强调了AI服务提供商有义务采取合理措施防止用户利用其平台侵犯他人著作权。此外，常熟市人民法院也判定AI生成的图像可以在一定条件下具有著作权[12]。这些案例表

9.参考Shuimei Liu, *Copyright Fair Use in the People's Republic of China--on the Road of Development: A Comparative Copyright Analysis of Chinese and the U.S. Fair Use, and Proposals for Corresponding Legislation in China* (Dec. 2021) (dissertation, Maurer School of Law - Indiana University).
10.Morgan Lewis, 北京法院认可人工智能生成图像的著作权保护, (2024年1月), https://www.morganlewis.com/blogs/sourcingatmorganlewis/2024/01/beijing-court-approves-copyright-protection-for-ai-generated-images.
11.Baker McKenzie,中国：关于AI生成作品著作权保护的里程碑式法院裁决, (2024年4月), https://insightplus.bakermckenzie.com/bm/intellectual-property/china-a-landmark-court-ruling-on-copyright-protection-for-ai-generated-works.
12.新华日报,江苏首例、全国第二例AIGC著作权侵权案件落槌引人深思——AI生成作品，谁才是真正的创作者，(2025年3月)。

明，中国法院在承认AI生成内容可能享有著作权的同时，也强调了不得侵犯现有著作权作品的原则。

在监管层面，中国也正在积极制定与AI和数据相关的法律法规和指导方针。中国国家互联网信息办公室（CAC）于2023年发布了关于生成式AI模型训练数据使用的指导意见，其中包括对使用受著作权保护信息的规定。同年生效的《生成式人工智能服务管理暂行办法》要求生成式AI服务提供者使用来源合法的数据和基础模型，尊重知识产权。此外，2024年发布的生成式AI服务管理条例草案提出了更详细的安全措施，包括对训练数据的安全评估，以及避免使用含有非法或有害内容的数据。这些监管举措表明，中国的监管部门正在努力构建一个既能促进AI发展，又能保障数据安全和知识产权的框架。

006>对中国AI行业的启示

"*Thomson Reuters v. Ross Intelligence*"案的判决结果对中国的AI行业，特别是那些开发法律科技或类似依赖大量数据进行训练的AI应用的企业，具有重要的警示意义。尽管中美两国的法律框架和具体判例有所不同，但著作权保护的基本原则是全球性的。

此案突显了即使是间接使用受著作权保护的材料（例如用于AI训练），也可能构成侵权，尤其是在开发与著作权所有者产品直接竞争的产品时。美国法院拒绝接受Ross公司关于其使用主题提要（head-notes）作为"中间步骤"的转换性使用抗辩，这警示中国企业在进行类似操作时需要格外谨慎。这意味着，仅仅因为AI的最终输出不直接复制训练数据，并不能保证训练过程本身不构成侵权。在涉及AI训练的案件中，法院可能会关注整个使用过程，而不仅仅是最终的输出。因此，中国AI企业在数据来源方面也面临风险。从互联网平台抓取包含受著作权保护材料的数据来训练AI模型，即使AI的输出不直接再现这些材料，也可能被视为侵权行为。

此外，中国AI企业还需注意，除了民事侵权风险，违反监管要求也可能带来行政风险。《生成式人工智能服务管理暂行办法》要求训练数据具有"合法来源"且不侵犯他人知识产权。如果企业的数据集包含未经

许可的受著作权保护内容，被监管机关发现或被投诉举报，可能被责令删除侵权数据、限期整改，严重者甚至会影响产品上线进度。因此，合规应当贯穿AI研发全流程。

007>合规建议

针对上述风险，中国AI企业在产品研发时应制定周详的数据合规策略：

- **优先使用合法公开数据或授权数据：** 尽可能使用公共领域或权利人许可的数据来训练模型。例如，法律领域可以利用公开发布的司法判决书原文（不受著作权保护的政府作品），避免使用他人整理的编辑性注释内容。又如训练通用语言模型，可优先选取政府公告、科学论文（作者授权开放访问）或Creative Commons协议许可的文本。对于受著作权保护的素材（文学作品、新闻稿等），应通过正式授权、购买著作权数据库等途径获取。

- **控制数据使用范围和比例：** 遵循最小必要原则，不超范围地收集受保护内容。如果出于算法效果需要引用少量他人作品片段，应确保使用量在合理限度内（如只占整体数据很小部分且不包含作品核心精华部分），并仅用于训练而非直接向用户提供原文。如果可能，尽量对数据进行去标识化和转换处理，例如提取特征而非保留可逆的全文，降低侵犯权益的风险。

- **建立内部审查和响应机制：** 企业应建立训练数据的审核流程，对数据来源、著作权状态进行标注分类。对于不确定是否侵权的素材，咨询法律专家意见或寻求权利人许可。保存数据获取和使用的记录，以备将来举证善意和技术用途之需。一旦收到权利人投诉，及时启动应对预案，必要时删除有争议的数据并调整模型，避免事态扩大。

- **关注法律和政策动向：** 持续关注国内著作权立法、司法实践对AI训练的态度。如果相关法律有新的例外规定或案例判决，应及时调整合规策略。例如，若未来中国引入"数据挖掘例外"规则，企业也需确保自身做法符合具体条件。同时，关注海外类似案件走向（如美国同类诉讼、欧盟文本数据挖掘指令等），因为中国企业的产品可能在海外市场

面临同样法律考验。

008 > 结语

综合来看，*Thomson Reuters v. Ross Intelligence*案体现了著作权法在AI时代对数据使用的红线：技术创新不能成为无视他人知识产权的借口。行业参与者在AI的开发和发展中应以此为鉴，在追求技术突破的同时，将法律合规作为重要考量。从长期看，只有尊重著作权、积极寻求授权合作，才能在避免法律风险的前提下安心拓展AI产品功能，推动产业健康发展。平衡好创新与著作权保护，中国的AI产业才能行稳致远。

（徐世达对本文亦有贡献）

顾萍
合伙人
知识产权部
纽约办公室
+86 10 5957 2089
guping@zhonglun.com

王成荫
非权益合伙人
知识产权部
北京办公室
+86 10 5780 8443
wangchengyin@zhonglun.com

伍波
非权益合伙人
知识产权部
北京办公室
+86 10 5957 2338
wubo@zhonglun.com

作 者 / 刘新宇

生成式AI的全链条数据合规要点及其风险防范

2025年1月，杭州深度求索公司发布的DeepSeek-R1模型在全球范围内引发广泛关注，其应用在短时间内登顶多国应用商店下载排行榜，表明我国生成式AI技术已具备国际竞争力。然而，技术的快速发展也带来了相应的数据合规挑战。从训练数据的来源合法性到生成内容的权利归属，从个人信息保护到跨境数据流动，生成式AI的全生命周期都面临着复杂的合规风险。

近年来，我国已初步建立起以《中华人民共和国网络安全法》（以下简称《网络安全法》）《中华人民共和国数据安全法》（以下简称《数据安全法》）《中华人民共和国个人信息保护法》（以下简称《个人信息保护法》）为基础，以《生成式人工智能服务管理暂行办法》为核心的人工智能监管法律框架。与此同时，欧盟《通用数据保护条例》（GDPR）等域外法规也对我国AI服务提供者的国际化发展产生了重要影响。在这种背景下，构建完善的全链条数据合规体系已成为生成式AI服务提供者可持续发展的必然要求。

本文旨在通过对现行法律法规、监管政策的梳理，分析生成式AI在数据合规方面面临的主要挑战，并提出相应的风险防范策略，为相关企业的数据合规提供参考。

001>生成式AI数据合规的法律框架

我国已构建了相对完善的数据合规法律体系，为生成式AI的合规发展提供了制度保障。这一体系以《网络安全法》《数据安全法》《个人信息保护法》三部基础性法律为核心，辅之以各类细化规定及国家标准，形成了多层次、全方位的监管框架。

其中，《网络安全法》确立了网络运营者的安全义务，要求其采取技术措施保障网络安全和数据安全。《数据安全法》建立了数据分级保护制度，明确了数据处理者的数据安全保护义务。《个人信息保护法》则从个人信息处理的全生命周期出发，规定了告知同意、最小必要等基本原则。这些法律共同构成了生成式AI数据合规的基础性规范。

而《生成式人工智能服务管理暂行办法》作为专门针对生成式AI的监管规定，确立了"包容审慎、分类分级"的监管原则。该办法从数据来源、内容生成、用户权益保护等多个维度构建了完整的合规框架。

002>生成式AI全生命周期所涉及的数据合规及规制

（一）数据收集合规

生成式AI技术的应用，以收集数据并进行加工、训练、处理为前提，因此数据来源的合法性是整个合规链条的基础。根据《生成式人工智能服务管理暂行办法》第七条的规定，生成式AI服务提供者必须使用具有合法来源的数据和基础模型，这一要求在实践中需从两个维度确保合规：1. 所获取数据的类型；2. 数据的获取方式。

1. 所获取数据的类型

在生成式人工智能数据治理框架下，数据类型的差异直接决定了差异化的合规义务体系。

首先，针对个人信息，处理个人信息需具备《个人信息保护法》第十三条规定的合法性基础。若以"取得个人同意"作为合法性基础，则需依据《个人信息保护法》第十七条规定，向个人告知信息处理者名称/姓名及联系方式、信息处理目的与方式、处理的信息种类、保存期限等事项。《生成式人工智能服务管理暂行办法》第七条进一步明确，服务提供者处理训练数据时若涉及个人信息，须依法取得个人同意或符合其他法定情形，同时应当遵循《网络安全法》《数据安全法》《个人信息保护法》中的相关要求，并采取严格的安全保护措施。

其次，针对敏感个人信息，法律设置了更为严格的合规门槛。根据《个人信息保护法》第二十八条、第二十九条规定，个人信息处理者仅在"具有特定目的、充分必要性且采取严格保护措施"的前提下，方可处理敏感个人信息；且处理时需取得用户单独同意，并满足法律法规对敏感信息处理的其他特别要求。以换脸AI应用采集人脸信息为例，若企业需将人脸信息作为AI模型训练的必要参数，需向用户作出特别告知并取得其单独授权同意。

再次，在重要数据和核心数据的处理方面，企业应依据《数据安全

法》及相关行业规定，开展数据安全风险评估，并实施与其重要程度相适应的特殊保护措施。

此外，公共数据的开发利用为生成式AI训练提供了重要来源，但其使用边界仍需谨慎把握。公共数据的使用应在合法公开的范围内进行，并遵循合理使用与可追溯原则。目前，上海、重庆等多地政府已建立起一体化公共数据平台，为AI训练提供了合法的数据来源，民间公开数据也在合规前提下被广泛运用。

2. 数据获取方式合规

从数据获取方式的角度考察，不同途径对应了不同的合规要求。

在直接收集用户数据的场景下，生成式AI服务提供者需要构建完整的告知同意机制。明确告知数据收集的类型、用途及保存期限等事项，且同意的获取需区分一般个人信息与敏感个人信息，敏感个人信息需取得用户的单独同意。同时建议生成式AI服务提供者完整留存用户的授权记录，留存期限应至少覆盖数据处理的全周期。

外部采购数据作为重要的数据获取渠道，其合规重点在于供应商管理与合同约束。虽然现行法律从促进数据要素流通的角度并未禁止商业性数据采购，但《生成式人工智能服务安全基本要求》等规范文件对商业语料采购提出了明确要求，包括但不限于：需具备法律效力的交易合同或合作协议、需对交易方/合作方提供的语料、承诺及材料进行合规审核。

网络爬虫作为高效获取数据的技术手段，其合规风险需要从多个维度进行系统评估。在技术层面，爬取手段的正当性涉及是否违反网站Robots协议、是否采用技术手段绕过访问限制、是否对目标网站的正常运行造成不当影响等关键问题。在内容层面，需要重点评估所爬取数据是否包含个人信息、商业秘密或受著作权保护的内容。在使用层面，则需警惕可能产生的不正当竞争风险，包括是否构成"实质性替代"、是否违反商业道德等。因爬取行为所涉及风险较高，AI服务提供者在向外部第三方爬取相关数据时，应妥善遵守关于爬虫抓取的合规要求，避免因此产生民事责任，甚至触发行政、刑事责任。

（二）数据处理及使用合规

数据处理是生成式AI技术落地的核心环节，其合规性直接关系到用户权益保护与技术创新边界的平衡。该阶段的合规风险主要聚焦于用户授权数据的二次利用。

从法律规范层面看，《个人信息保护法》第二十三条明确规定，个人信息处理目的、处理方式发生变更的，应当重新取得个人同意。这意味着，若生成式AI服务提供者最初收集数据的目的为"即时响应用户需求"，后续拟将其用于"模型优化训练"，必须履行单独告知义务并获取新的授权。GDPR亦规定，若个人信息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相关情况，且"二次利用"的目的应与初始收集目的兼容。

二次利用的潜在风险集中于数据泄露与权益滥用。当用户数据进入训练数据库后，存在被模型"记忆"并在对其他用户的响应中被模型无意披露的风险，既可能泄露个人隐私，也可能侵犯商业秘密。正是基于此原因，国外多家知名企业要求员工谨慎使用ChatGPT等生成式AI产品。2024年4月，我国支付清算协会也发布相关倡议，要求从业人员谨慎使用此类工具，避免数据跨境泄露风险。

（三）数据内容与质量合规

在生成式AI的监管体系中，数据治理对数据内容的合法性、安全性及数据质量的规范性亦提出明确要求，而数据清洗作为保障数据内容与质量的核心手段，既是法律法规的要求，也是提升生成式AI服务质量的内在需求。

从法律层面看，《生成式人工智能服务管理暂行办法》构建了数据内容与质量管控的基础框架。其第四条明确生成式AI服务需遵循核心原则，包括禁止生成危害国家安全、传播暴力色情等违法有害内容，防范民族、性别等维度歧视，尊重知识产权与他人合法权益，并提升服务透明度及生成内容准确性。第七条则针对训练数据处理活动进行细化规定，强调数据与基础模型来源合法，保护知识产权与个人信息，同时通过有效措施提升训练数据的真实性、准确性、客观性及多样性，且需符合《网络安全法》《数据安全法》《个人信息保护法》等法律法规要求。

数据清洗的实施亦有助于生成式AI服务质量提升。在合规标准方面，《生成式人工智能服务安全基本要求》通过附录A负面清单，明确需规避的违规内容范畴（如违反核心价值观、歧视性内容、侵犯权益内容等），为企业自查纠偏提供参照。在实施路径上，建议通过"机器过滤+人工审核+投诉响应"的多元机制开展数据清洗，重点聚焦个人信息保护与知识产权领域以规避侵权风险，同时需按法规要求开展数据标注及质量评估，确保语料的准确性与安全性。

（四）数据存储合规

生成式AI的数据存储合规，核心关注点为数据存储地点与存储期限的规范性，同时需提升数据泄露风险的应对能力。在存储地点方面，企业应优先选择数据本地存储，非必要不进行跨境存储，以规避跨境存储及访问可能引发的多区域规制冲突。在存储期限方面，需依据"最小必要原则"收集并处理含个人信息在内的各类数据，并结合行业属性、数据类型及相关法律法规要求，确定数据的最短必要存储期限，避免因长期存储带来合规风险。

此外，企业还需建立数据泄露风险防控方案与应急预案，以应对生成式AI应用过程中潜在的数据安全事件。例如，ChatGPT在企业应用场景中曾多次发生数据泄露事件，此类案例为行业敲响了警钟。

（五）数据跨境合规

数据存储的地域选择进一步引出了生成式AI的数据跨境合规问题，除一般数据与个人信息跨境传输需遵循的授权同意规则外，企业还需重点警惕潜在的数据跨境风险。实践中，生成式AI运营涉及的数据跨境情形主要包括两类：一是数据出境，如境内企业将服务器部署于境外导致训练信息向境外传输，或采购、嵌入境外生成式AI产品/服务进行加工训练；二是数据入境，如境内企业使用境外语料训练自有生成式AI产品，或境外生成式AI产品将生成结果传输至境内。

当前各国均对数据出境（而非数据入境）严格规制，数据若涉及跨境传输，主要适用数据出境一方所在地的监管法律法规。例如，2025年1月28日，意大利数据监管部门收到针对DeepSeek的投诉，指控其

违反欧盟及本国数据保护法规，核心问题包括：Deepseek作为共同数据控制者未在欧盟设立机构，亦未按GDPR第27条要求指定欧盟代表；其隐私政策虽表明用户数据存储于中国，但未提及标准合同条款、约束性公司规则等跨境传输保护措施，也未公开数据传输影响评估文件。同时，隐私政策在存储期限、用户权利行使方式、未成年人数据保护等方面存在信息不完整或表述模糊的问题，不符合GDPR相关条款要求。最终意大利监管部门于2025年1月30日在该地区禁用DeepSeek。

从这一案例可看出，开展跨境生成式AI服务的企业，除满足常规合规要求外，还需充分符合数据跨境传输规则，严格遵循数据出境一方所在地的监管政策，避免因跨境合规疏漏引发监管处罚。

003>结语

在生成式AI快速发展的背景下，数据作为核心生产要素，其全生命周期的合规管理已成为AI服务提供者保障业务合法运营、规避监管风险的关键。生成式AI服务提供者落实数据合规需聚焦五大核心环节：在数据来源端，按数据类型履行告知同意等合法性义务，确保语料获取渠道合规；在数据处理端，以充分授权为前提，严格限定使用范围，超出原范围时需重新获取同意；在数据质量端，通过数据清洗确保语料合法、真实；在数据存储端，强化安全保障措施与企业数据安全体系建设；在数据跨境端，先识别传输必要性与豁免情形，再依法履行出境安全评估、标准合同备案等义务，并符合境外法规要求。

这些措施相互衔接，共同推动生成式AI的数据合规建设。这不仅是企业实现稳健发展的必然选择，也能为生成式AI产业规范创新奠定基础，助力行业在合规框架下释放技术价值。

刘新宇
合伙人
知识产权部
上海办公室
+86 21 6061 3666
jeffreyliu@zhonglun.com

作者 / 蔡鹏

AIGC语境下的版权保护边界初探

1.1 AIGC的“黑箱”特性对传统版权理论的冲击

生成式人工智能（Generative AI）及人工智能生成内容（AI-Generated Content, AIGC）相关技术，特别是其“黑箱性”（black box）与“不可预测性”（unpredictability），正对传统版权法理论构成根本性挑战。传统版权理论假定：创作者——无论是作家、画家还是作曲家——对其创作工具（如笔、画刷或乐器）拥有近乎完全的物理控制，从而将其内心的“思想”精确地“表达”为最终的作品。

然而，AIGC打破了这一“思想—工具—表达”的三角关系。当用户输入一个提示词（prompt）时，即便是高度详细的指令，AI模型如何选择内容、如何组织像素、如何呈现风格，对人类用户而言在很大程度上是不可预见和难以精确控制的。这种“控制”的缺失，使得“创作”行为本身的法律判定变得明显复杂化，从而突出了传统版权理论在应对AIGC时的制度困境。

1.2 人类在何种程度上贡献了AIGC作品的独创性

AI本身能否成为作者？对于这一问题很容易形成共识。版权法保护的是人类的作品，并非机器。法律分析的焦点已经转向一个更具争议性且亟待解决的前沿问题：如果AI被定位为“作者的画笔或照相机”，即一种高级的创作工具，那么使用该工具的“人”，需要做出何种程度、何种性质的贡献，才能被法律认定为该AIGC作品的“作者”？这一问题的核心，在于法律审查的重心必须从评估“AI的自动化程度”转向“人类所贡献的独特产出”。

1.3 保护有质量的“智力体现”是核心

在美国版权局（United States Copyright Office，USCO）拒绝*Théâtre D'opéra Spatial*一案的登记申请中，艺术家Jason Allen声称其输入了“至少624个提示词”并进行了“大量修改”，但这一巨大的“物理贡献量”并未使其获得作者身份。此案生动地揭示了AIGC的版权保护边界不应锚定在人类贡献的“最终篇幅”或“物理数量”（如修改次数、

提示词字数）上，更不应锚定在人类对AI"黑箱"的"机械控制"上。

相反，版权保护的核心锚点应当是"**人类在作品中的智力体现并予以表达**"。这包括创作者独特的构思与表达、作者与表达的对应性，以及在生成过程中体现其主观选择的迭代筛选与表达修改。

本文旨在通过中美比较法的路径，初步系统地论证为何这种"智力体现"的质量（而非其数量或控制）是界定AIGC版权唯一合法且有效的标准。

002>比较法的基础：中美对"人类作者"原则的坚守与演绎

在AIGC版权问题的全球讨论中，中美两国作为关键的司法管辖区，在法律的基石——"作者"必须是"人"——这一原则上，达成了高度共识。然而，两国达成共识的路径与法理演绎却不尽相同，而正是这种演绎的差异，为后续认定标准的分歧埋下了伏笔。

2.1 美国法下的"基石"：*Thaler v. Perlmutter* 案的系统性论证

美国法下对"人类作者"原则的坚守，在*Thaler v. Perlmutter*一案中得到了系统、权威的阐述。该案中，申请人斯蒂芬•泰勒(Stephen Thaler)博士将其AI系统"Creativity Machine"列为一幅名为《通往天堂的新近入口》（A Recent Entrance to Paradise）的视觉艺术作品的唯一作者，并主张AI应被承认为作者。

USCO拒绝了该申请，理由是作品缺乏"人类作者"。经过逐级上诉，美国哥伦比亚特区联邦巡回上诉法院在2025年3月作出最终裁决，明确指出"**人类作者身份是美国版权法的基石性要求**"（Human authorship is a bedrock requirement of copyright）。

法院指出，《版权法》中的诸多核心制度设计，均预设了"作者"必须是"人类"：

其一，财产权主体：版权是一种财产权，其主体必须是能够拥有财产的法律实体，而机器不具备此资格；

其二，保护期限：版权保护期通常以"作者的有生之年加70年"计算，而机器没有"寿命"；

其三，继承权：版权涉及配偶、子女等人类家庭关系，机器没有继承人；

其四，意图（intention）：合作作品要求作者之间存在"共同创作意图"，而机器没有意识，无法形成法律意义上的"意图"。

2.2 中国法下的共识：AI作为"画笔"或"照相机"的工具定位

与美国通过标志性判例进行体系化法理论证不同，中国司法实践通过个案中的"工具论"比喻，生动阐述了中国《著作权法》中"作者必须是人"的基本原则。

在北京互联网法院审理的"春风送来了温柔"案（中国首例明确支持AI生成图片构成作品的判决）中，法院明确将AI模型Stable Diffusion比作"**作者的画笔或照相机**"。法院认为，AI是作者用于创作的辅助工具，其法律地位与传统的创作工具无异。

在其他相关判决中，我国法院维护了《著作权法》的主体必须是自然人的法律原则。这种将AI作为"工具"的定位，将法律审查的焦点从AI本身，转移到了使用AI的"人"身上。

2.3 从原则到实践：共识所遗留的核心问题

当中美两国都确认AI只是"工具"，而"人"才是潜在的作者时，一个远比"AI是否是作者"更棘手、更复杂的法律问题浮出水面：使用照相机作为工具的人，按下快门即可在绝大多数情况下被认定为作者；但使用AIGC作为工具的人，输入一个简单的提示词（如苏州"蝴蝶椅子案"）时，却可能不是作者。

人类与AI合作产生作品，到底需要做什么才能成为法律意义上的作者？核心问题就在传统的版权"思想与表达二分法"，并据此重新划定人类贡献的准入门槛。

003>共同的边界："思想与表达二分法"下的提示词（prompts）定性

"思想与表达二分法"是版权法的基本原则。法律只保护思想的独创性表达（如小说的情节设定、美术作品的构图色彩），而不保护抽象的构思、主题、方法等"思想"本身。

在AIGC场景中，人类用户最直接的贡献——提示词（prompts），在法律上应如何定性？在这一点上，中美两国的司法和行政实践再次达成了共识：**简单、常规的提示词本身属于"思想"范畴，不受版权保护。**

3.1 中国司法实践：上海"提示词"案

中国首例对AI提示词本身是否构成作品进行裁判的案件——2025年上海市黄浦区人民法院就某案件作出的判决，为提示词的法律定性提供了清晰的参照。

该案中，原告主张其为AI绘画平台Midjourney撰写的一系列"精心设计"的提示词，构成了受保护的文字作品。法院经审理后，驳回了原告的全部诉讼请求，其核心理由有二：

其一，缺乏独创性。法院认为，尽管原告的提示词看似复杂，但其内容"各元素间仅为简单罗列，缺乏语法逻辑关联""关键词组无序组合"。更重要的是，这些提示词"缺乏作者的个性化特征"，所选用的艺术风格、材质等均属"该领域常规表达"，未能体现作者独特的审美视角或艺术判断。

其二，属于思想范畴。法院进一步运用"思想与表达二分法"明确指出，涉案提示词"仅体现抽象的创作想法和指令集合"，其核心是"对画面元素、艺术风格、呈现形式等的罗列与描述"，这些内容更多属于"抽象的创作构思，属于思想范畴"。

本案判决为"思想"设定了一个明确的下限——即便是经过选择和堆砌的、描述性的"指令集合"，如果缺乏"个性化特征"和"作者特有的逻辑"（"表达"的结构），就仍然停留在"思想"层面，不受版权法保护。

3.2 美国行政实践：USCO的立场

USCO在其发布的《版权登记指南：包含AI生成材料的作品》（2023年）及《版权与人工智能报告》（2025年）中，采取了与上述上海法院实质相同的立场。

USCO明确指出，提示词（prompts）在功能上"更像是对委托艺术家的指示"（instructions to a commissioned artist），其本身是"不受保护的想法"（unprotectible ideas）。

USCO论述了提示词与"控制"标准的关系：即使用户提供了"高度详细的提示词"，用户也无法"控制AI系统如何处理这些元素"，以及如何将其"表达"出来。因此，提示词只是"想法"，而AI（机器）完成了最终的"表达"。在*Théâtre D'opéra Spatial*案中，USCO审查委员会也重申，无论提示词多么复杂，其本质都是在"接受"AI系统对指令的"解释"，而非作者自身的表达。

3.3 比较法视野下的共识与张力

中美两国在AIGC版权审查的起点上达成了一致：简单的、常规的、作为"指令集合"的提示词，其本身（无论作为文字作品还是作为创作贡献的证据）均位于"思想"一侧，被排除在版权保护之外。

然而，这种共识的表面下隐藏着明显的分歧。上海"提示词"案否定的是"缺乏个性化"的提示词；USCO否定的则是"无法控制"的提示词。这留下了一个共同的、悬而未决的关键问题：如果一个提示词（或一系列提示词的迭代过程）既体现了"远超常规的独特构思与个性化表达"，又能通过参数设置和迭代修改（如"春风案"）对AI施加了足够的"智力投入"而非"机械控制"，它是否还能被轻易地归为"思想"？

在这一点上，中美的司法实践开始分道扬镳。

在如何认定"人"的贡献何时能跨越"思想"门槛、构成受保护的"独创性表达"这一核心问题上，中美两国则展现出了根本分歧。美国采取了"表达元素的实质性创作控制"标准，而中国则探索了"过程导向的智力投入"标准。

4.1 美国标准："实质性创作控制"的严格审查

美国版权局和法院的审查核心是"人类创作控制"。USCO的指南明确要求，AIGC作品要获得保护，人类作者必须对最终作品的"表达元素"（如线条、色彩、构图）拥有"足够的"或"实质性的"控制。在当前技术下，USCO普遍认为，仅仅输入提示词"不能提供足够的控制"，因为用户无法"控制"AI如何"表达"该提示词（思想）。

这一严格标准在以下两个标志性案例中得到了充分体现。

其一是*Zarya of the Dawn*案。该案中，艺术家Kristina Kashtanova使用AI工具Midjourney为其漫画书生成插图。USCO在审查后，对这部漫画作品进行了"机械性拆分"处理：

拒绝保护图像：USCO撤销了对由Midjourney生成的单张图片的版权保护，理由是Kashtanova虽然输入了提示词，但她无法"预测"Midjourney的输出结果，因此她对这些图像缺乏足够的"创作性控制"。

批准保护汇编：USCO保留了对整部漫画书的版权登记，保护范围被严格限定为Kashtanova自己创作的文本内容，以及其对文本和图片的"选择、协调和编排"。

该案较完整地展现了美国"控制"标准的严格性。USCO将一部有机的漫画作品强行"分离"为了AI生成的图像（因人类无控制，故无版权）和人类编辑的汇编（因人类有控制，故有版权）。

其二是*Théâtre D'opéra Spatial*案。如果说*Zarya of the Dawn*案是"严格拆分"的范例，那么*Théâtre D'opéra Spatial*案则展现了"控制"标准的严格性。该案中，艺术家Jason Allen使用Midjourney生成的作品获得了科罗拉多州博览会的艺术奖项。在向USCO申诉时，Jason Allen强调了他对上述作品付出的"巨大的人类努力"，包括为了实现其"特定的艺术

愿景"而进行的"至少624次提示词和迭代"。

然而，USCO审查委员会坚决拒绝了Jason Allen的登记申请，认定上述作品"缺乏人类作者身份"，核心理由是：当AI根据提示生成复杂作品时，其"传统创作要素"是由技术（AI）自主决定的，用户无法行使"最终的创造性控制"。

此案是美国"控制"标准基于传统版权理论的典型应用。即使Allen的"特定艺术愿景"（即"独特构思、设定、风格"）是清晰的，他也为此付出了"624次提示词和迭代"的巨大努力，但USCO仍然认为：当人类无法控制AI时，"智力体现"在法律上或者至少在版权无法获得保护。

上述案例体现了美国的"实质性创作控制"标准，特别是其对"表达的控制性"的刚性要求，这在AIGC场景下显得较为严苛。该标准所要求的"控制"在AIGC的场景下很难实现（尽管需要逐案评估）[1]。该标准最终奖励的是AI生成之后的修改和编辑（如Zarya of the Dawn案的"汇编"），而非AI生成前和生成中的"独特构思"。

4.2 中国标准："过程性智力投入"的灵活探索

与美国严格的"控制"标准形成鲜明对比，中国法院在司法实践中探索了一条更为灵活、务实的路径，即"过程导向的智力投入"标准。

该标准的审查重点不在于人类是否对最终输出的"每一个像素"拥有控制力，而在于"人类用户在AI生成过程中的具体行为"。法院会综合考察用户是否通过输入提示词、调整参数、选择模型、后期修改等一系列行为，对最终生成的内容施加了足够的"智力投入"和"个性化表达"。

这一标准在以下两个标志性案例中得到了充分体现。

其一是"春风送来了温柔"案。北京互联网法院在该案中支持了原告对AI生成图片的著作权主张，其判决理由的核心在于：**法院不强求**原告能够完全"控制"AI工具Stable Diffusion的输出，而是关注原告在生成图片过程中的行为，即原告"**通过多次调整提示词、设置相关参数**"，对画面元素、布局构图等"表达细节"进行"**选择和安排**"。法院认为，这一个

1.US Copyright Office Releases Part 2 of AI Report，https://www.akingump.com/en/insights/ai-law-and-regulation-tracker/us-copyright-office-releases-part-2-of-ai-report .

性化的过程，清晰地体现了原告的"**审美标准和个性判断**"。因此，最终的图片（作为"表达"）体现了原告的"**独创性智力投入**"。本案中，法院认识到，人类的"独特构思"（审美标准、个性判断）可以通过"风格指令及修改"（调整提示词、设置参数）这一过程，被有效地注入到最终的AIGC作品中。

其二是上海"提示词"案与苏州"蝴蝶椅子"案。"智力投入"标准并非意味着无条件保护所有AIGC作品，上海"提示词"案和苏州"蝴蝶椅子案"作为反向判例，证明了该标准的有效性和对称性。

在苏州"蝴蝶椅子案"（全国首例否认AI文生图可版权性的案件）中，法院否定了涉案AI生成图片的独创性，理由是原告输入的提示词"属于相对简单的叠加……对画面元素、布局构图等描述缺少差异性"，且被告举证证明在原告之前已有类似概念的作品出现。同样，在上海"提示词"案中，法院也认定提示词"缺乏作者的个性化特征……属该领域常规表达"。

并置分析上述三地法院的判决，可以清晰地看到我国的司法裁判，已经为"智力投入"标准建立了一个相对宽松的"标尺"：**当人类给予了独特、个性化、高投入的"智力体现"时**，其作品（或贡献）能够跨越"思想"的门槛，构成受保护的"表达"；**而当人类仅给予常规、简单、缺乏差异性的"智力体现"时**，其贡献则停留在"思想"层面，不受保护。

笔者认为，中国司法实践中的"过程性智力投入"标准，在一定程度上比美国的"实质性创作控制"标准更具灵活性和前瞻性。它不强求人类对AI"黑箱"的物理控制，而是关注**人类心智活动（审美、判断、选择）的质量**。从法院的思路上看，它鼓励创作者进行"深度的人机协作"，利用AI作为"画笔"来实现其"独特构思"，而非因遵循严格的"控制"标准放弃对人机作品的保护。

但是，这种标准也存在过于宽松的问题。如何体现投入高？多次修改提示词是否就等同于投入高？如何对此进行理解和限缩，才能更符合《著作权法》"激励创作、促进科学与艺术的进步"的立法本意？

005>传统"思想—表达二分法"在AI时代承受压力

"思想—表达二分法"是《著作权法》中最核心、最基础的法律原则，也是判定侵权与否的"黄金标尺"。简而言之，《著作权法》只保护思想的"表达"，而不保护"思想"本身。随着全球化，这一原则被写入《与贸易有关的知识产权协定》（TRIPs）第9条第2款："版权保护应延及表达，而不延及思想、程序、操作方法或数学概念本身。"中国作为WTO成员国，在《著作权法》及相关司法实践中也应严格遵循这一原则。

法律之所以如此小心翼翼地剥离思想与表达，是基于公共利益与个人激励的平衡。如果"思想"可以被垄断，文化的源头就会枯竭，其中蕴含的基础逻辑是：版权法只赋予作者对自己"劳动成果"（具体的表达）的垄断，而不允许其圈占"人类智慧的公有领域"（思想）。但是，在AI时代，生成式人工智能以及大模型技术本身所抓取、分析、利用的，不正是人类的所有智慧？其数据和算法相结合所带来的"涌现"，不也正是"人类智慧"？从这个角度来说，笔者认为，"思想—表达二分法"本身没有过时，但在AI时代如何理解，则成为了一个新的问题。

AIGC场景对传统版权领域的挑战，在于如何有效解释"思想—表达二分法"；尤其是随着AI技术的进步，提示词本身趋于简单化、同质化，而最终的作品则是对作者从不同的思考角度对AI工具进行反复提问，并经由人机协作后不断修订和完善的体现。以文字作品为例，尽管最终作品的80%有可能为机器生成，但是作品本身的逻辑、框架以及最终的"华彩"部分，均为作者自己完成。

正如上述部分法院判决将AI比作"画笔"或者"照相机"，笔者认为，**AIGC的创作过程是一种"人类对机器的驯化"。**AI模型是充满随机性和海量数据训练的"野马"，而人类的独特智力体现（通过提示词、参数、后期修改体现）则是"缰绳"。最终的作品是人类的意志和审美驯化AI的随机性后产生的固定表达。

此时，法律不应再问"你是否100%控制了马的每一步？"，而是"**最终马跑出的路线，是否体现了你的意志和技巧？**"

006>结论：如何重塑AIGC版权的保护边界

笔者认为，AIGC 版权保护的边界，应且必须锚定于人类在创作过程中注入的"符合原创标准的智力体现与表达"上。这种"体现与表达"应被定义为：

一是构思的独特性，即超越"常规表达"的结构设定、角色定义或艺术构思。

二是指令的个性化，即超越"简单罗列"的、具有"个性化特征"的风格指令、参数设置和体现了"个性判断"的原创选择。

三是过程的迭代性，即超越"一次性输入"的、可证明的"多次调整"，以及经过"筛选、修改、编排"的、体现作者独立创作的过程。

如果上面的定义过于冗杂，以下则为更加直接的结论：

首先，简单的提示词绝对属于"思想"，不受保护。

其次，复杂的"结构化提示词"，如果本身包含了一段独创的故事、一段诗歌，或者极其独特的各种形容词的排列组合，那么提示词文本本身可能作为"文字作品"受保护，但这并不意味着通过其生成的图片/结果自动受到保护。

最后，能通过证据（图片、文本的修订记录）证明作者本人对AI工具最终生成内容的"原创性表达"施加了决定性影响的情况下，人类即成为该作品的合法作者，其著作权应受法律保护。

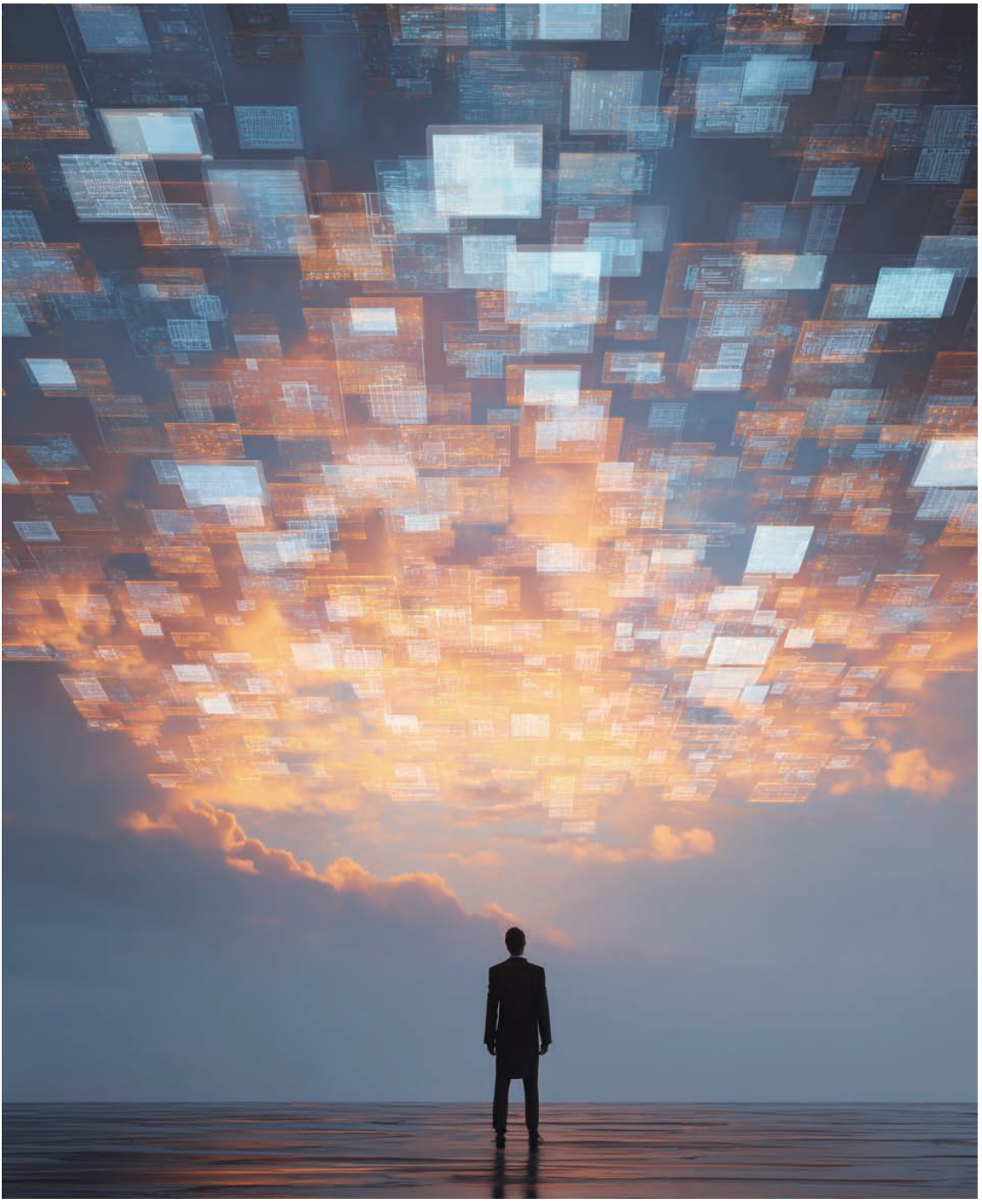
笔者预测，未来，人机共创将会成为所有人文艺术领域的主流形态。以文字作品为例，未来的作者将会把主要的时间和精力放在构想本身，而大部分具体的语言组织可能都会由机器"代劳"。当具体的"遣词造句"由机器代劳时，判定作品"原创性"的重心将不可避免地发生从"表达层面"向"选择与安排层面"的迁移。

蔡鹏
合伙人
知识产权部
北京办公室
+86 10 5087 2786
caipeng@zhonglun.com

chapter 04

人工智能之产业治理|热点观察

artificial intelligence industry governance : hot topics observation



作 者 / 王飞 贺梦琳

AI智能体的法律问题透视

2025年6月，被视为首个真正自主的人工智能体Manus推出全新的Chat模式，并对所有用户免费开放，用户通过登录获取积分可以免费体验Manus的AI对话功能。而值得注意的是，在支持即时问答的同时，用户可无缝链接Manus Agent模式，即Manus可直接执行对话中所涉及的复杂任务，例如问题研究、文档制作、网页设计、代码编程或数据分析。AI智能体（AI Agent）将AIGC应用实现从"回答问题"向"解决问题"模式进行转换，这也是基础大模型能力升级的必然结果。此前笔者在《AIGC产品的生命周期透视（上）（下）》[1]系列文章中系统分析了AIGC产品数据、代码、大模型等各环节可能面临的法律问题，而AI智能体作为AIGC的进化功能在AIGC产品的基础上将涉及其特有的法律问题，本文将对此进行针对性讨论。

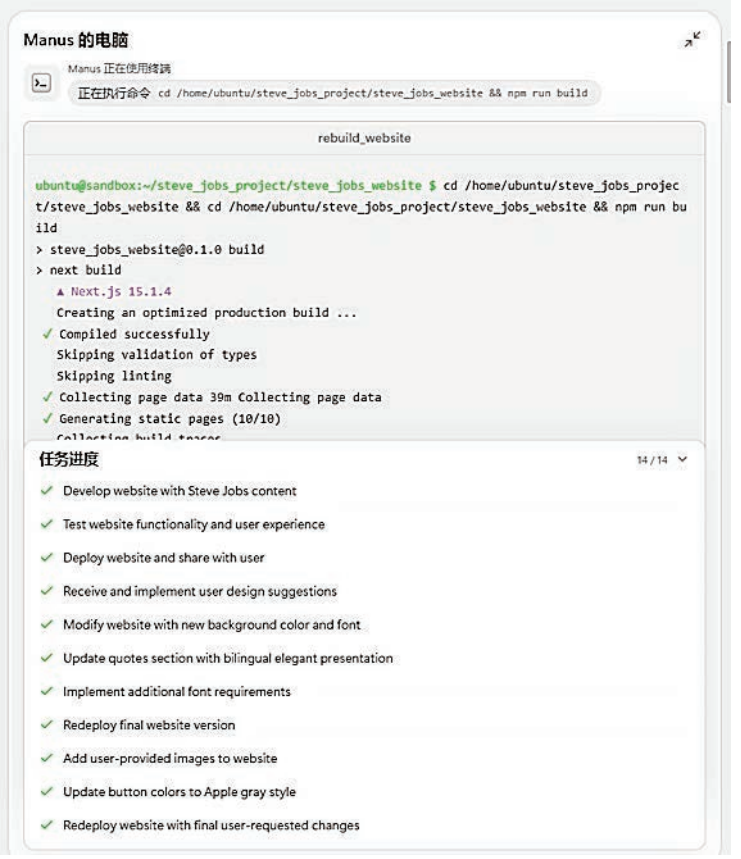
001> AI Agent的技术概念

Agent，即智能体，一般认为它是一种能够感知所处环境，并依据所感知到的信息自主做出决策并执行相应行动，以实现特定目标的实体。这一概念涵盖了软件、硬件以及虚拟系统等多种形式。AI Agent训练已经可以实现在现实世界中进行多模态理解的能力。它为利用生成式人工智能和多个数据源进行训练提供了一个框架。[2]而Anthropic则从另一个角度对智能体进行定义，其指出智能体体现了自主任务导向的工作逻辑：AI Agent由LLM动态地指导其自身流程和工具使用，即智能体在根据任务要求决定执行步骤、使用工具和流程控制上具有极高的自主性。

以Manus为例，"Manus"在拉丁文中意为"手"，象征着知识不仅存在于思维中，还应能通过行动得以实现。这体现了Agent与AI Bot（聊天机器人）产品从回答问题到执行任务的本质进阶。Manus具备执行人类所能完成的智力任务的能力，不仅可以为用户提供想法，更重要的是能将想法付诸实践，真正解决问题。[3]从技术角度来看，Manus的核

1.https://mp.weixin.qq.com/s/asKSyvtdTm2OcJO3uYUDkw.
https://mp.weixin.qq.com/s/kAUzbnA7qXH72ITo-tpcOg.
2.李飞飞等：Agent AI：Surveying the Horizons of Multimodal InterAction.
3.Manus百度百科：https://baike.baidu.com/item/Manus/65463546.

心工作原理是实例化一台远程的虚拟机沙箱环境，后续所执行的命令、代码均在沙箱环境中运行，在整个会话结束前会一直保留，在这过程中模型可以随时创建目录、读取文件，能做到信息的存储和交互等。[4]具体执行步骤是：第一，依赖于大语言模型的视觉能力获取信息；根据关键词信息调用第三方API，获取搜索结果。如果已经在页面中获得需要分析的关键数据，不会再进行模拟点击。如果没有获得关键信息，则会进入到二级页面进行模拟浏览器点击网页内容、浏览网页内容以及获取网页文本内容。Manus可基于页面的文字进行爬取，获取文本形式的视觉信息。第二，通过代码运行python等程序，创建excel、ppt等文件对数据进行分析和处理。第三，利用大语言模型的视觉能力，通过部署静态前端页面向用户展示成果。在Manus提供的"乔布斯人物分析与网站制作"的示例任务交互页面显示，Manus由虚拟电脑执行了完成任务所需数据检索、分析、开发、测试和生成的流程。



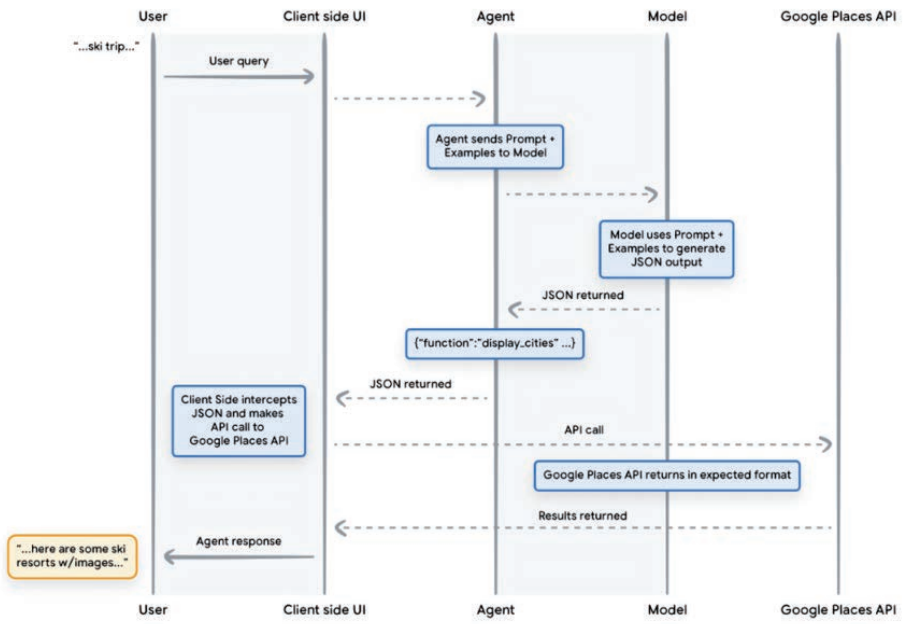
Manus官网提供"乔布斯人物分析与网站制作"示例交互页面

4.阿里云开发者，《Manus的技术实现原理浅析与简单复刻》，https://mp.weixin.qq.com/s/SSO-w6FF4mBm2zrXY5RzkA.

继Manus以后，智能体已经成为AI公司投资和研发的重点，ChatGPT于2025年2月推出深度思考（Deep Research）功能，可以根据用户指令输出复杂问题的研究报告；Anthropic于2025年5月推出Claude 4模型，专注于提升AI执行复杂任务的能力，定位于任务执行引擎和AI Agent系统的核心组件；2025年4月，字节跳动推出扣子空间通用型AI Agent产品，定位于用户的协同办公场所；2025年5月，昆仑万维天工超级智能体APP上线，定位于全球首款基于AI Agent架构的Office智能体手机APP。

002 > AI Agent特有的法律问题之数据获取合规性

Google Agents白皮书指出，与大模型所获知的知识仅限于训练数据中包含的内容不同（非联网状态下），智能体还可通过工具接入外部系统获取拓展知识。Agents根据设定的自定义代码或拓展工具，可以处理用户输入的查询需求以获取相关信息并调用API，在Agents无法实时调用API（如时间或操作顺序限制）或调用未暴露到互联网的API时，则可通过在客户端应用程序调用函数管理API执行。



Google Agents白皮书绘制Agent应用中函数调用生命周期序列图

正如美国已有批量的版权人针对AI未经授权使用其作品用于训练数据提起诉讼（相关案例可参考此前笔者《以全球范围AIGC诉讼为例梳理AIGC的侵权认定和权利限制规则》[5]一文），AI Agent自动调用API以获取相应数据或信息将可能面临更严峻的合规问题。在我国司法实践中，互联网平台对依法搜集的数据享有合法权益，非法调用API获取相应数据并进行商业性使用有违公平、诚信原则和商业道德，扰乱了数据市场竞争秩序，涉嫌构成不正当竞争行为。[6]Agent具有极高的自主性，在没有用户明确指示的前提下在其虚拟机中自动执行链接API的操作，如果链接未向互联网公开的API接口，可能涉及绕过数据防护措施、违反robots协议对网页内容进行爬取；通常而言，违反robots协议、破坏或绕过技术保护措施、大量多次异常爬取网页内容的行为均可能面临不正当竞争风险。当然，实践中利用非法爬取的数据是否进行了实质性替代的使用、是否破坏产业生态和竞争秩序也是认定不正当竞争行为的重要考量因素。而AI Agent向用户输出的结果具有任务专一性，其目的通常不是为了实质性替代网页提供的服务，而是对多网页提供的内容进行加工、分析，由此，从实质性替代角度分析，是否应直接将Agent通过算法自主调取API行为定性为不正当竞争行为有待商榷。

除涉及不正当竞争风险外，智能体还可能自动链接API直接调用他人的作品进而涉及著作权侵权风险。近日美国法院在Meta案和Anthropic案中，暂认定AI训练数据使用受版权保护的书籍可构成合理使用。这也与美国在2025年5月美国版权局发布《版权与人工智能报告第三部分：生成式人工智能训练》发表观点相一致，其认为政府干预训练行为为时过早，合理使用制度可以根据个案情况、多因素平衡决定。而我国《著作权法》第二十四条并没有为人工智能训练数据的行为提供直接的权利限制，正如笔者在《以全球范围AIGC训练数据侵权诉讼为例梳理合理使用规则的适用》[7]一文中所述，AI训练数据行为的法律性质目前仍需在三步检验法的框架下个案确定，而AI Agent调用API使用他

5.https://mp.weixin.qq.com/s/gnmYhJFheLTJNgI3YksPkw.
6.广东省高级人民法院，广东高院终审宣判首例涉数据抓取交易不正当竞争纠纷案，https://mp.weixin.qq.com/s/sD2FQEp-OCH6ZfDxRzOPJug.
7.https://mp.weixin.qq.com/s/uI_13p0CVtrScomZ9XDO4w.

人作品的行为是否可能构成转换性使用而予以免责也将有抗辩空间（当然以生成内容不与原作品构成实质性相似为前提）。

正因AI智能体能够独立从各种来源收集和处理数据，其运行的过程具有一定的"自主性"，决策过程隐藏在复杂的算法和不断演化的数据集中，导致信息来源缺乏透明度，而平台通常会进行数据来源澄清声明，以避免可能的侵权争议。例如OpenAI声明其基础模型主要使用三个主要信息来源：（1）互联网上公开可用的信息；（2）与第三方合作的信息；（3）OpenAI用户、训练人员和研究人员提供或生成的信息。对于公开可获取的互联网信息，OpenAI宣称只使用可在互联网上自由和公开获取的信息，不会故意获取付费才可获取的信息或从暗网收集数据。[8]Manus在提供的任务示例中也展示了其调用数据的来源，通常为公开社交媒体和网页的数据，如下图所示。

Manus官网提供"乔布斯人物分析与网站制作"示例交互页面

笔者认为，仅使用互联网公开的信息的确可以降低侵权风险，但随着智能体输出内容的执行程度越来越高，垂直领域智能体必然在AI Agent市场占据一定地位，而对于垂直领域智能体而言，提高输出内容的权威性和可信度才是智能体产品真正的市场竞争力，由此，与第三方专业数据库合作从而保证输出高质量内容将可能是更优解。

8.https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-foundation-models-are-developed.

003 > AI Agent特有的法律问题之生成内容可版权性

AIGC版权归属问题本身具有一定争议，美国版权局一直拒绝对AI生成作品进行版权登记，理由包括绝对的AI自动生成内容由机器生成，而只有人类创作的作品才能受到版权法保护，以及人类对AI生成内容缺乏创作性的控制（输入相同的提示词，生成内容却大相径庭）等。美国版权局在2023年3月准许《玫瑰之脸》（Rose Enigma）作品版权登记，但登记范围仅限于"申请材料中清晰可见且可与申请人放弃权利主张之外的、与非人类表达相区分的、未经修改的人类绘画作品"。2025年1月，美国版权局准许《一块美国奶酪》（A Single Piece of American Cheese）作品版权登记，但登记范围仅限于构成完整图片的元素（即AI生成的部分材料）的"排列组合"（可理解为汇编作品）。目前，中国法院在AIGC的版权认定上主要考虑用户的独创性智力投入因素，在"春风送来了温柔"案中，北京互联网法院认定用户通过AI工具四轮修改形成的图片具有审美意义，具备智力成果和独创性要件，可以认定为作品。[9]但在"蝴蝶椅子"案中，张家港市人民法院认为"对于主要由人工智能绘图软件自动生成的内容，不应当认定构成作品……用户首次输入提示词，即可生成体现提示词主题和要素的图形，这其中人工智能对文字到图形的生成，仍然起到重要作用，尚不能体现对图形的充分独创性"，在本案中AIGC生成内容没有被认定为作品。

笔者认为，主流国家普遍认为机器无法成为人而具有版权法意义上作者的身份，如果用户对智能体输出的成果主张权利，考虑到人类对相关内容生成的控制程度进一步降低，由此相较于AIGC产品而言，AI Agent生成内容可被版权法保护的可能性更低。例如用户要求Manus生成"乔布斯人物分析与网站制作"，其自主执行了搜索、浏览、创建文件、生成并测试网站的过程，生成的网站中包括自主选择的图片和文

9.最高人民法院，因为一张AI生成的图片，他告到了法院！一系列问题摆在法官面前……https://mp.weixin.qq.com/s/Nad-PYQewzsIIDLcKoGUn9g.

字以及子页面设计，用户可以在此基础上要求Manus对生成的网站进行修改，在这其中用户对AI Agent成果的贡献程度较低，如适用美国版权局的逻辑，则用户在生成AI Agent成果的过程中没有体现人类表达、选择和编排，可能较难被认定为版权法意义上的作品。但实际上，智能体生成成果的价值却相较于通用AIGC更高，用户使用智能体的个性化投入也不一定完全减少，由此，智能体成果的可版权性判断标准需要考虑更多因素，例如智能体为不同用户生成相同成果的概率、智能体生成侵权的任务成果的概率、AI智能体的工具属性等。

004 > AI Agent特有的法律问题之侵权责任承担

AI Agent具有高度的自主性，相较于传统的人工智能而言，人类对于生成结果的可控性进一步降低。尽管它可能会提供非常简单的过程以解释结果的生成，但隐藏在其结论背后的产生过程仍然是算法运行的结果。这种自动决策过程使得结果具有高度的不确定性，即使AI Agent生成了侵权的内容，被侵权一方往往难以确定错误的具体来源，从而使得寻求补救或分配责任变得复杂。在人工智能主要被视为工具的时代，责任归属通常较为明确——使用工具的个人或实体需对其行为负责。然而，以Manus为代表的AI Agent打破了这种简单的责任归属认定。如何在用户与开发者划分责任的归属，需要回归到现有法律框架和责任划分体系进行探讨。平台作为AI Agent的重要载体，对于AI Agent生成结果侵犯他人权利的问题，AI Agent平台是否应当承担责任，应当承担何种类型的责任，其注意义务的限度如何界定等是亟待解决的关键问题。

从当前的司法案例来看，现有的两例“人工智能平台侵权案”为我们提供了参考。首先，要根据Agent平台实施的具体行为区分侵权责任的类型。例如在（2024）浙0192民初xx号案（以下简称“**浙江案**”）中，法院对平台的身份进行了认定。法院认为某智能科技有限公司通过某AI平台应用生成式人工智能技术向用户提供在线AI图像生

成、LoRA模型训练等生成式人工智能服务，因此，某智能科技有限公司属于生成式人工智能服务提供者。平台作为运营者，并未参与到用户实施的上传参考图片、发布和分享生成图片的行为之中；从主观方面来看，亦缺乏证据证明平台在用户的涉案行为中，与用户之间存在共同提供作品的意思联络，因此，平台不是网络传播内容的提供者，未直接实施受信息网络传播权控制的行为，不构成直接侵权，应当认定构成信息网络传播权的帮助侵权行为。而在（2024）粤0192民初xx号案件中（以下简称"**广州案**"），法院认为平台的行为构成直接侵权。平台通过用户提供的关键词为用户提供侵权内容，意味着平台提供的模型可以直接生成与案涉角色实质性相似的图片，构成对原作品复制权和改编权的直接侵权。基于前述案例可知，要在个案中具体考量AI Agent平台实施的具体行为，以区分平台应当承担的责任类型。其若直接生成了侵权的内容，则可能构成直接侵权；若侵权行为系用户实施，平台作为人工智能服务提供者，若未尽到合理的注意义务（难以以避风港原则进行抗辩），可能要承担帮助侵权的责任。

如何界定AI Agent平台的注意义务范围是另一个关键问题。区别于一般的网络服务提供者的注意义务（"通知-删除规则"），对于AI Agent而言，平台注意义务的界定应综合考量AI Agent的服务性质、发展水平、被诉侵权事实的明显程度、平台营利模式、可能引发的侵权后果、可以采取的必要措施及其效果、侵权责任承担对行业的影响等因素，将平台注意义务确定在与其信息管理能力相适应的程度。例如在"杭州案"中，法院首次明确指出，生成式AI服务提供者兼具"平台管理者"与"内容生产者"双重身份。平台作为生成式人工智能服务提供者，不能仅仅依据避风港原则进行抗辩。法院认为被告虽然已经采取关键词过滤等措施，停止生成相关图片，并达到了一定效果。然而，当向平台输入与案涉角色相关的其他关键词时，仍可产生实质性相似的图片。因此，被告应进一步采取关键词过滤等措施，防范其服务继续生成与案涉角色作品实质性相似的图片，防范程度应达到：用户正常使用与案涉角色相关的提示词，不能生成与案涉角色作品实质性相似的图片。

更进一步的，在"广州案"中，由于平台构成直接侵权，法院明确平台的注意义务边界应以《生成式人工智能管理暂行办法》对于生成式人工智能服务提供者的要求进行判断，包括但不限于是否建立了投诉举报机制、是否进行了潜在风险提示以及是否进行了显著标识等。

综上所述，需要在实践中进一步考察AI Agent应用的发展阶段以及平台的管理能力，追溯AI Agent生成侵权内容的来源，查明平台在其中的参与度，以此为基础界定其注意义务，以便合理分配各方主体责任。

005 > AI Agent特有的法律问题之法律主体资格

即使AI Agent具有高度的自主性，其在法律上的地位仍然是"物"或"工具"等客体。有观点认为，如果一个AI Agent能自主评估交易风险、与对方Agent进行谈判，进行合同签署，实施法律行为等，我们是否需要考虑在特定、严格限定的领域内，承认它具有某种"缔约代理资格"？换言之，可否因为AI Agent的高度自主化而承认其在一定程度下作为"法律主体"而存在？

基于人类的主体价值的哲学视角，若赋予非人类主体以法律上的主体资格，需要有充分的理由。以民法当中的"法人"制度为例，法人在民法中也是民事主体，是为了实现让其承担责任、降低交易的成本和风险的目的，本质上仍然体现了人的目的性。《民法典》中对于法人制度的拟制有着充分的理由，很显然，在AI Agent发展的起步阶段，其难以类比法人制度而拥有法律主体的资格。即使要探讨其在特定场景下的"有限法律角色"，也要充分考量实践中Agent的发展程度、技术水平，考量其是否具有享有权利或承担责任的功能等，也不能损害人类的主体性地位，不扰乱现有法律秩序的稳定。

现阶段，基于"人的主体价值"的基本哲学理念，以及AI Agent发展的现状，赋予其法律主体资格面临重大挑战。因此，在坚持"工具论"的前提下，对于AI Agent完成的指令或决策、生成的内容，该如何进行法律上的定性以及责任的承担是需要进一步研究的问题。

“知识产权法是技术之子”，以Manus为代表的AI Agent爆火，为AI进一步发展赋能，实现了AI技术再次飞跃，但还需调整法律制度以适应新时代下技术的变革，以法律制度为技术发展“保驾护航”。

（赵好对本文亦有贡献）

王飞
非权益合伙人
争议解决部
北京办公室
+86 10 5087 2877
philipwang@zhonglun.com

作 者 / 李瑞 马悦 蒲昱含

人工智能生成合成内容标识合规十问十答

随着人工智能技术的飞速发展，AI生成合成内容已经深入到我们生活的方方面面。从文本创作到图像生成，从音频合成到视频制作，AI技术正在重塑内容创作和传播的方式。然而，技术的进步也带来了新的挑战：如何区分人工智能生成内容与人类创作内容？如何防范AI技术被恶意利用？如何保护公民、法人和其他组织的合法权益？

2025年3月14日，国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局联合发布了《人工智能生成合成内容标识办法》（以下简称《标识办法》）。此外，国家市场监督管理总局、国家标准化管理委员会还并配套发布了强制性国家标准GB 45438-2025《网络安全技术 人工智能生成合成内容标识方法》（以下简称《标识方法》）。随着上述规定于2025年9月1日起正式生效施行，一套覆盖全链条、多模态的AI内容标识合规体系已然确立。近期，网信部门集中查处一批存在人工智能生成合成内容标识违法违规问题的移动互联网应用程序，依法依规予以约谈、责令限期改正、下架下线等处置处罚，有关人工智能生成合成内容标识的合规监管已然落地。本文将以问答的形式，就相关企业在履行标识义务过程中常见的问题进行释明，帮助企业搭建其标识合规体系。

001> 为什么要对人工智能生成合成内容进行标识？

《标识办法》第1条明确其立法目的为"促进人工智能健康发展，规范人工智能生成合成内容标识，保护公民、法人和其他组织合法权益，维护社会公共利益。"具体来说，在当前人工智能技术迅猛发展的背景下，AI生成合成内容的真实感不断增强，部分内容已使得普通网络用户难以辨别其所接触信息是否来源于AI技术。此类内容若被不当利用、过度滥用乃至恶意操纵，将可能引发显著的社会安全与信任风险。

标识制度并非要限制AI技术的发展，而是要为技术发展划定边界、建立规则。一方面，对人工智能生成合成内容进行标识是保障公众知情权与信息真实性的关键举措，能够让受众清晰辨别信息来源，防止因误信虚假内容而遭受误导、欺诈，也因此有助于避免AI合成内容被不当利

用，保护公民、法人和其他组织合法权益，维护社会公共利益；另一方面，通过明确的标识要求，可以提升AI技术安全水平，增强公众对AI技术的信任，为技术的健康发展和商业化应用创造良好的社会环境。

002 > 提供人工智能生成合成内容服务的平台应当如何履行标识义务？

作为人工智能生成合成内容的主要源头，提供人工智能生成合成内容服务的平台承担着重要的标识义务。根据《标识办法》的规定，服务提供者应当履行以下义务：

（1）**开展标识活动**：服务提供者需要根据《标识办法》第4条和第5条的规定对生成合成内容添加显式标识，并在生成合成内容的文件元数据中添加隐式标识；服务提供者提供生成合成内容下载、复制、导出等功能时，应当确保文件中含有满足要求的显式标识；

（2）**在用户协议中进行说明提示**：在用户服务协议中明确说明生成合成内容标识的方法、样式等规范内容，并提示用户仔细阅读并理解相关的标识管理要求；

（3）**依用户申请提供不添加显式标识的内容**：用户申请服务提供者提供没有添加显式标识的生成合成内容的，服务提供者可以在通过用户协议明确用户的标识义务和使用责任后，提供不含显式标识的生成合成内容，并依法留存提供对象信息等相关日志不少于六个月；

（4）**其他协助义务**：在履行算法备案、安全评估等手续时，应当提供生成合成内容标识相关材料，并加强标识信息共享，为防范打击相关违法犯罪活动提供支持和帮助。

003 > 如果提供的产品/服务没有包含人工智能生成合成内容的功能，是否还需要关注标识义务？

《标识办法》不仅规定了人工智能生成合成内容服务提供者的标识义务，还规定网络信息内容传播服务提供者、互联网应用程序分发平台等主体也负有对自身平台上传播、分发的生成合成内容进行标识的法定义务。

实践中，如果企业提供的产品/服务没有包含人工智能生成合成内容的功能，还应关注其是否因构成生成合成内容的分发、传播主体而需要履行相应的标识义务。具体来说：

（1）**互联网应用程序分发平台**：在应用程序上架或者上线审核时，应当要求互联网应用程序服务提供者说明是否提供人工智能生成合成服务。互联网应用程序服务提供者提供人工智能生成合成服务的，互联网应用程序分发平台应当核验其生成合成内容标识相关材料；

（2）**网络信息内容传播服务提供者**：如果企业提供的产品/服务允许用户上传内容，即使产品本身没有AI生成合成功能，也可能需要承担传播服务提供者的标识义务，包括：

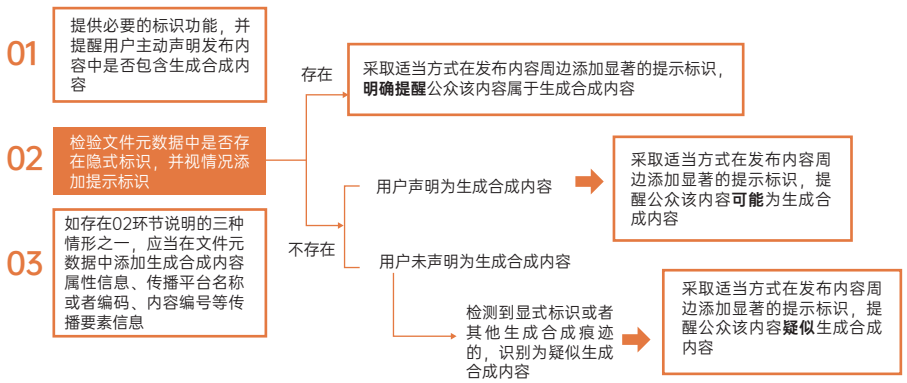


图 1 网络信息内容传播服务提供者的标识义务

004 > 传播平台可否通过要求用户不上传人工智能生成合成内容，来避免承担标识义务？

仅通过用户协议或其他方式要求用户不上传人工智能生成合成内容，并不能完全免除传播平台的标识义务。在实践中，考虑到现有的人工智能生成合成内容检测技术往往不够成熟和完善，无法100%准确识别所有人工智能生成合成内容，加之海量内容使得辅以人工审核的成本极高，用户还可能通过技术手段去除或篡改标识，完全禁止用户上传人工智能生成合成内容几乎是不可能的；在此背景下，《标识办法》第6条所规定的传播服务提供者的标识义务（具体见上文第3问所述）是强制性的，我们倾向于认为企业作为传播平台不能通过用户协议或其他方式要求用户不上传人工

智能生成合成内容来避免履行以上标识义务。要想妥善管理相关合规风险，还是需要妥善履行第3问中介绍的传播平台的相关标识义务。

005 > 人工智能生成合成内容标识里的显式标识和隐式标识分别长什么样？

显式标识是在人工智能生成合成内容或交互场景界面中添加的，以文字、声音、图形等方式呈现并**可被用户明显感知**到的标识。例如，文本内容显式标识是在文本的起始、末尾或者中间适当位置添加的文字提示或者通用符号提示等标识，或者在交互场景界面、文字周边添加的显著的提示标识；图片内容显式标识是在图片的适当位置（边或角）添加的显著的提示标识；交互场景界面显式标识是在内容附近持续显示的提示文字或者在交互场景界面顶部、底部、背景等适当位置持续显示的提示文字。就标识内容而言，无论标识的对象为何种形式（文本、图片、音频、视频、虚拟场景、交互场景界面），显式标识都应当同时包含人工智能要素和生成合成要素，即：包含"人工智能"或"AI"，表明使用人工智能技术；以及包含"生成"和/或"合成"，表明内容制作方式为"生成"和/或"合成"。就标识形式而言，根据《标识方法》的规定，针对不同的标识内容通常会采取包括文字、角标、语音标识、音频节奏标识等不同的标识形式。以下是部分显式标识的示例：

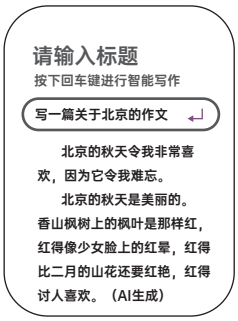


图2：文本内容显式标识示例

图3：图片内容显式标识示例

隐式标识是采取技术措施在人工智能生成合成内容文件数据中添加的，不易被用户明显感知到的标识，主要包括文件元数据隐式标识

和内容隐式标识。其中，文件元数据隐式标识包含生成合成标签要素、生成合成服务提供者要素、内容制作编号要素、内容传播服务提供者要素以及内容传播编号要素；在格式方面，文件元数据隐式标识应符合《标识方法》附录E的规定。以下是部分隐式标识的示例：

图片元数据字段

compression	:	JPEG
thumbnail_offset	:	205926
thumbnail_length	:	15963
XMP toolkit	:	XMP Core 4.4.0-Exiv2
AIGC	:	{ "Label" : "1", "ContentProducer" : "15652010000000016BDWXYY0400000000", "ProduceID" : "v0300fg100kbc3c77ub10123456", "ReserveCode1" : "e862483430d978cbf828b8b24296ef9328d843a0", "Propagator" : "1561101000000057WMWB030000000000", "PropatorID" : "qdn57u6mld93z6o1xvz", "ReserveCode2" : "e862483430d978cbf828b8b24296ef9328d843a0"}
profile_CMM_type	:	
profile_version	:	0.0.0
profile_class	:	display device profile

图4：图片文件元数据隐式标识示例

006 > 显式标识和隐式标识是不是二选一即可？

根据《标识办法》的规定，显式标识和隐式标识不是替代关系，而是并存互补关系。首先，就其功能定位来看，显式标识主要用途是向公众提示内容由人工智能生成合成，防止公众被误导或欺骗；隐式标识主要用途是记录生成合成内容相关信息，便于监管部门追踪内容来源，为侵权追溯提供技术支持。一般而言，人工智能生成合成内容服务提供者需要同时对生成合成内容添加显式标识和在生成合成内容的文件元数据中添加隐式标识；但在部分特殊情形下，如服务提供者提供的生成合成内容不存在文件元数据，在技术上无法实现对文件元数据的隐式标识，则不需要添加隐式标识，但依然需要对生成合成内容添加显式标识。

007 > 如果提供的产品/服务不是文本、图片，而是音频、视频、虚拟场景，如何进行标识？

《标识办法》和《标识方法》对于音频、视频、虚拟场景等的标识要

求亦进行了明确规定。具体来说，在标识内容方面，如上文第5问所述，无论标识的对象为何种形式（音频、视频、虚拟场景），其显式标识都应当同时包含人工智能要素和生成合成要素；在标识形式、标识位置等方面，《标识方法》对于音频内容、视频内容和虚拟场景的显式标识分别进行了规定。

就文件元数据隐式标识而言，音视频内容的文件元数据隐式标识在内容和格式方面与文本、图片的文件元数据隐式标识所遵循的要求实质一致，具体见上文第5问的说明。以下是部分音视频生成合成内容的显式标识和隐式标识示例：

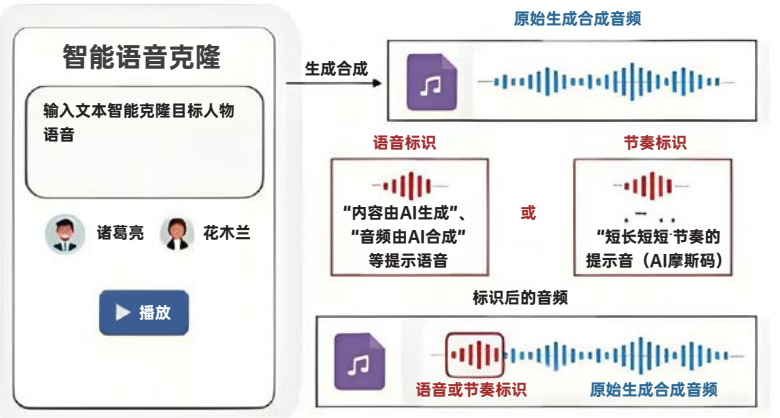


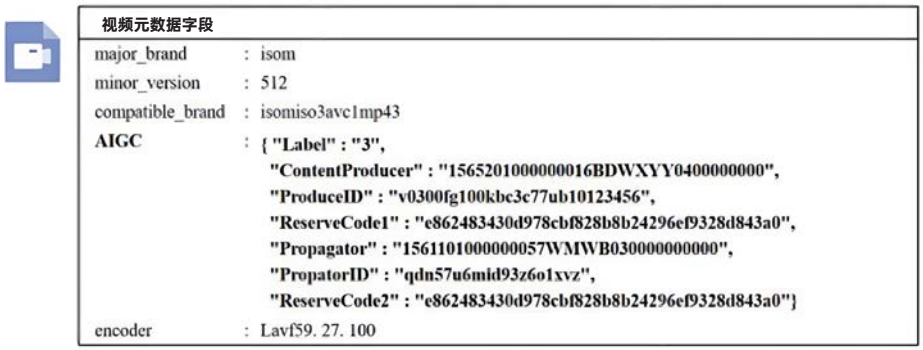
图5：音频内容显式标识示例

图6：视频内容显式标识示例

音频元数据字段

major_brand	: M4A
minor_version	: 512
compatible_brand	: isomiso4
AIGC	: { "Label" : "2", "ContentProducer" : "15652010000000016BDWXYY0400000000", "ProduceID" : "v0300fg100kbc3c77ub10123456", "ReserveCode1" : "e862483430d978cbf828b8b24296ef9328d843a0" , "Propagator" : "1561101000000057WMWB030000000000", "PropatorID" : "qdn57u6mld93z6o1xvz", "ReserveCode2" : "e862483430d978cbf828b8b24296ef9328d843a0"}
encoder	: Lave59. 27. 101
duration	: 00:04:55, 58
start	: 0.025057
bitrate	: 128kb/s
stream#0	: 0
audio	: mp3, 44100Hz, stereo, s16p, 128kb/s

图7：音频文件元数据隐式标识示例



视频元数据字段

major_brand	: isom
minor_version	: 512
compatible_brand	: isomiso3avc1mp43
AIGC	: { "Label" : "3", "ContentProducer" : "1565201000000016BDWXYY0400000000", "ProduceID" : "v0300fg100kbc3c77ub10123456", "ReserveCode1" : "e862483430d978cbf828b8b24296ef9328d843a0", "Propagator" : "1561101000000057WMWB030000000000", "PropatorID" : "qdn57u6mid93z6o1xvz", "ReserveCode2" : "e862483430d978cbf828b8b24296ef9328d843a0"}
encoder	: Lavf59. 27. 100

图8：视频文件元数据隐式标识示例

008 > 如果违反标识相关规定，会面临哪些法律责任和风险？

根据《标识办法》的规定，“违反本办法规定的，由网信、电信、公安和广播电视等有关主管部门依据职责，按照有关法律、行政法规、部门规章的规定予以处理”。此处的有关法律法规包括但不限于《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》等。违反上述规定，可能会面临包括警告，通报批评，责令限期改正，责令暂停提供相关服务，责令暂停信息更新并处罚款，治安管理处罚，甚至追究刑事责任等在内的法律责任和风险。

009 > 出海企业在海外是否也要关注人工智能生成合成内容标识合规要求？

目前，生成合成内容标识义务正在逐渐成为一项国际"惯例"，不少海外国家已经或正在探索将标识义务纳入法律法规之中。出海企业在海外展业时同样应当关注人工智能生成合成内容标识的合规要求，降低企业在当地的违法违规风险。以下是一些典型海外国家的法规要求：

- **欧盟**发布的《人工智能法》（*the AI Act*）第50条要求生成合成音频、图像、视频或文本内容的人工智能系统（包括通用人工智能系统）提供者，应确保人工智能系统的输出内容以机器可读格式进行标识，且可检测其为人为生成或操纵（artificially generated or manipulated），并应在技术可行的前提下保证这种标识的技术方案具备有效性、可互操作、稳健性以及可靠性。生成或操纵构成"深度伪造"的图像、音频或视频内容的人工智能系统部署者，应披露该内容为人为生成或操纵。

- **美国**加州2024年8月发布的AB-3211《数字内容溯源标识法案（草案）》曾提议在人工智能生成内容中嵌入可溯源数据以确保内容来源的透明度，加州州长2024年9月19日正式签署的《加州人工智能透明度法》（*the California AI Transparency Act*）明确要求每月有超过100万名用户的生成式人工智能系统提供者应提供免费的AI检测工具，允许用户验证内容是否由AI生成，并履行针对人工智能生成内容提供显性标识功能以及设置隐式标识的义务。

- **越南**的《数字技术产业法（草案）》也提出企业应以机器可读和可检测的格式标记人工智能系统的输出。

010 > 总结：面对人工智能生成合成内容标识要求，企业应该采取哪些合规措施？

面对人工智能生成合成内容标识新规的实施与落地，我们建议企业应采取系统性的合规措施，以确保全面满足法律法规要求。具体来说：

- **高度重视，积极应对**。人工智能生成合成内容标识不是可选项，

而是必选项。企业应当从战略高度认识其重要性，投入必要资源以确保合规。

- **系统规划，分步实施**。标识合规是一项系统工程，需要从管理、法律、技术等多个维度统筹规划，循序渐进地推进实施。在管理方面，我们建议企业结合最新法律要求，为开展标识工作制定相应的工作方案/制度，并明确各部门的职责分工，确保各部门在开展标识工作过程中有效协作；在法律方面，我们建议企业加强内部有关标识新规的培训和学习，实时跟进监管部门对新规的解读和最新监管实践；在技术方面，我们建议企业开发或采购符合法律要求的标识技术，及时了解技术发展动态，不断改进标识技术。
- **专业支持，风险可控**。就开展标识工作过程中遇到的具体疑问和难点，我们建议企业寻求专业的法律和技术支持，获得专业的法律指导和合规建议，从而确保所采取合规措施的科学性和有效性，将合规风险降至最低。
- **持续改进，与时俱进**。考虑到AI技术和相关法规都在快速发展，我们建议企业应建立动态调整机制，实时关注最新法规要求，加强与同行的交流，总结经验与最佳实践，持续优化合规体系。

李瑞
合伙人
公司业务部
北京办公室
+86 10 5957 2143
lirui@zhonglun.com

作者 / 蔡鹏

GDPR 处罚下的 AI 出海合规：核心隐私风险、典型案例与治理路径

引 言 > 1500 万欧元罚单——OpenAI 案敲响的警钟

2024年12月20日，人工智能企业出海的合规叙事被彻底改写。意大利数据保护机构（Garante）针对OpenAI的数据处理活动违反《通用数据保护条例》(GDPR) 作出公开决定，处以1500万欧元罚款。这一裁决不仅是欧盟对生成式人工智能 (GenAI) 巨头的首个重磅罚单，更是对全球AI行业监管"教育期"结束的明确信号。

Garante认定的违规行为直指AI数据处理的全生命周期：

- **数据泄露通知义务：**未就2023年3月的数据泄露事件正确通知监管机构（违反 GDPR 第33条）。
- **合法性基础：**缺乏将用户个人数据用于模型训练的合法性基础（违反 GDPR 第 6 条）。
- **未成年人保护：**对未成年人保护不足（违反GDPR第8条、第25条）。
- **透明度义务：**违反信息义务，隐私政策不完善（违反GDPR第12条、第13条）。
- **监管命令：**未按要求履行信息宣传义务(违反GDPR第83条)。

据报道，这笔罚款基于OpenAI 2023年度全球总营业额的1.58% 核算。它暴露了即便是行业领袖，在GDPR严格的监管框架下也存在系统性的合规短板。因此，深入剖析此案，是理解AI企业在欧盟市场生存风险的第一步。

001 > OpenAI 案的核心教训——"合规后补"模式的彻底失败

Garante对OpenAI的处罚，其意义远超罚款金额。它在法律逻辑上对科技行业普遍采用的"业务先行，合规后补" (launch first, patch later) 模式进行了根本性否定。

Garante的监管底牌："问责制原则"

Garante的处罚逻辑并非仅仅针对违规的事实，更是针对OpenAI无法证明其合规的状态。Garante严密论证的核心武器是GDPR第5条

第2款，即"问责制原则"（Accountability Principle）。

该原则要求数据控制者（OpenAI）必须能够证明其落实了数据保护的各项原则。证实拥有合法性基础的责任在于OpenAI自身，监管机构仅需对OpenAI提供的举证进行有效反驳。

Garante正是抓住了时间线上的致命缺陷：

1.**业务先行：**OpenAI的ChatGPT服务于2022年11月30日推出。

2.**合规后补：**OpenAI提交的合规证据——DPIA（数据保护影响评估）报告于2023年2月24日编写；LIA（正当利益评估）报告初稿同日完成。

Garante强调，数据处理的合法性基础必须在处理活动开始前便成立。OpenAI提供的文件晚于服务上线时间，**因此**，Garante认定OpenAI未能证明其在服务推出之前就已完成适当性评估和合法性基础的确定。基于此，Garante裁定OpenAI在2023年3月14日更新隐私政策之前，"缺乏为模型训练处理用户数据的合法性基础"。

"业务先行，合规后补"模式的终结

Garante的这一裁决逻辑，是对AI行业依赖快速迭代和大规模数据处理模式的正面打击。在传统模式下，DPIA（数据保护影响评估）/TIA（传输影响评估）等合规文件常被视为业务上线后的"待办事项"，在业务逻辑验证后才被动完成。

但Garante的立场表明：没有事前的、可证明的合规文档，数据处理活动从一开始就具有"原罪"。这种"合规原罪"无法通过后续的补救措施（如补写LIA报告）来追溯豁免。

因此，对于企业高管和法务而言，这传递了一个清晰的信号：隐私合规（特别是GDPR第25条所规定的"设计和默认的数据保护"）必须嵌入产品研发(R&D)的初始阶段，而不是法务部门在产品发布前的最后一道关卡。

下表清晰地展示了OpenAI的系统性合规失败及其法律依据：

表 1：Garante 对 OpenAI 核心违规裁定与法律依据

违法行为	违反的GDPR条款	Garante的核心论据
1.数据泄露报告义务错位	第 33 条	泄露时爱尔兰公司未成立，不适用"一站式监管"，应分别通报包括意大利 Garante 在内的所有相关监管机构。
2.缺乏合法性基础	第6条, 第5条第2款	无法提供证据，证明在服务上线前（2022.11.30）已确定用于模型训练的合法性基础。
3.隐私政策（信息义务）	第12条, 第13条, 第5条第1款(a)项	仅提供英文；访问路径不佳（注册后才能阅读）；对非用户（训练数据）未告知；对用户数据训练的描述过于笼统。
4.未成年人保护不足	第8条, 第24条, 第25条	未设置年龄验证机制；无法确保与未成年人签订的服务条款（合同）在意大利法律下的有效性。
5.未履行信息宣传命令	第83条第5款(e)项	未按114/2023号决定要求，有效触达公众开展信息宣传活动，未能告知用户其权利。
6.生成内容准确性（移交DPC）	第5条第1款(d)项	Garante：透明度（提示不准）不能豁免遵守准确性原则（提供准确信息）的义务。

002 > "Garante 模式"——从"事后罚款"到"业务阻断"的新型执法武器

分析OpenAI案，如果只看到2024年的1500万欧元罚款，就忽略了其在2023年遭遇的更致命打击——"业务阻断"。

GDPR：GenAI的"事实监管者"

2023年GenAI呈爆炸性普及时，欧盟专门的《人工智能法案》(*AI Act*) 尚在立法进程中。此时，欧盟各成员国的数据保护局 (DPAs) 并未等待新法，而是迅速部署了GDPR这一在设计上"技术中立"的法规。

DPAs利用GDPR实质上成为了全球GenAI的"事实监管者"(de facto regulators)。在这场监管浪潮中，意大利Garante扮演了"领头羊"的角色，其行动揭示了一种比传统罚款更有效的新执法策略。

"以市场准入为筹码"的新策略

2023年3月30日，Garante宣布对ChatGPT发布"立即执行的临时限制"（即临时禁令），导致其在意大利市场即刻无法访问。这一行动是监管博弈的分水岭。

传统上，GDPR的跨境执法（特别是一站式机制）可能需要数年才能作出罚款决定。但Garante利用了GDPR第66条的"紧急程序"权力，在极短时间内绕开了漫长的调查流程，直接切断了OpenAI的业务。

这种"以市场准入为筹码"的"Garante模式"，其即时威慑力远超事后罚款。它带来的影响是：

1.**风险升级**：合规风险从"财务成本"（CFO的问题）升级为"业务存亡"（CEO的问题）。

2.**迅速整改**：OpenAI在不到一个月的时间内（2023年4月28日）即完成了Garante要求的重大整改，以换取其在意大利市场的服务恢复。

3.**全球效应**：Garante迫使OpenAI推出的整改措施，包括"全球适用的年龄验证机制"和数据训练"选择退出"(Opt-out)功能，其影响远超意大利国界，实质上提升了全球AI产品的合规基线。

因此，Garante的行动证明，"业务阻断"比"事后罚款"具有更强大的即时威慑力。AI企业出海面临的首要威胁已从滞后的财务处罚转变为即时的市场准入封锁。

003 > DeepSeek案警示——"零容忍"与不可触碰的数据主权红线

如果说2023年的OpenAI案是Garante的"教育与谈判"，那么2025年的DeepSeek案则展示了监管的"零容忍与迅速阻断"。

执法策略的演变：从"教育与谈判"到"零容忍"

执法策略的演变清晰地体现在时间线上：

- **对OpenAI(2023)**：Garante给予了近一个月的补救窗口期。
- **对DeepSeek (2025)**：2025年1月，DeepSeek AI应用在欧洲下载量激增。Garante于1月28日发出信息请求，短短两天后（1月30日）即宣布在意大利封禁 (Block) DeepSeek AI。

这种执法策略的急速转变表明，欧盟DPA对GenAI的"宽限期"已经结束。对于在OpenAI案后仍未吸取教训、在基础合规上"完全不合格"的新进入者，监管机构将不再提供谈判机会。DeepSeek案迅速引发了多米诺骨牌效应，比利时、法国、德国的DPA均已确认启动跟进调查。

DeepSeek触碰的"致命红线"：非法国际数据传输

Garante采取迅速封禁的核心原因，在于DeepSeek触及了GDPR的根本红线，特别是具有高度地缘政治敏感性的数据主权问题 。

1. 非法的国际数据传输 (违反GDPR第44条)

这是"最致命的问题"。DeepSeek的早期隐私政策公开声明，收集的欧盟用户数据将存储在中华人民共和国 (PRC) 。

- **法律背景**：欧盟与中国之间没有"数据保护充分性认定"。
- **合规要求**：因此，此类传输必须依赖 GDPR第46条规定的严格保障措施（如标准合同条款,SCCs），并辅以复杂的传输影响评估 (TIA) 。
- **违规事实**：DeepSeek显然未能提供任何此类保障，这被Garante视为"不可接受的数据主权风险" 。

2. 公然无视 GDPR

DeepSeek的隐私政策甚至没有提及GDPR，这被视为对欧盟法律的"公然漠视" 。

3. 不合作态度

Garante称DeepSeek对质询的回复"完全不充分"，甚至有报道称其辩称无需回应当地监管机构 。

DeepSeek案表明，对于非欧盟/非充分性认定国家的AI企业（特别是中国企业），数据传输合规 (GDPR Chapter V) 已成为一个决定生死的"红线"。这不是一个可以罚款了事的合规项，而是一个决定市场准入

的"准入问题"。

下表汇总了近期欧盟DPA的重点执法行动，清晰地展示了监管的趋势：范围在扩大（从聊天机器人到社交媒体 AI），执法在提速（从月到天），力度在加强（从谈判到封禁）。

表 2：欧盟 DPA 对 GenAI 重点执法行动对比 (2023-2025)

目标公司（产品）	主要DPA	日期	指控的核心GDPR违规行为	结果 / 状态
Luka Inc. (Replika)	Garante (意)	2023年2月	第8条(对未成年人的风险)，缺乏年龄验证	处理禁令
OpenAI (ChatGPT)	Garante (意)	2023年3月	第 6 条 (合 法 性),第 8 条 (未成年人),第5条(准确性)	临时禁令 (后恢复)
OpenAI (ChatGPT)	EDPB (欧盟)	2023年4月	协调各DPA对第6,5条等问题的调查	成立ChatGPT工作组
X (Grok)	DPC (爱尔兰)	2025年	第6条(使用"公开帖子"训练AI 的法律依据)	正式调查中
DeepSeek (D e e p - Seek AI)	Garante (意)	2025年1月	第 4 4 条 (非 法 数据传输至中国)，第13/14条 (透明度)	服务被封禁
DeepSeek (D e e p - Seek AI)	比利时/法国/德国 DPA	2025年1-2月	跟进数据传输、GDPR合规性指控	多国启动调查

004 > AI出海面临的四大系统性合规风险

综合OpenAI、DeepSeek、Replika 和Grok案，可以发现AI企业出海面临的合规风险并非孤立事件，而是呈现出四大系统性、根本性的规律和逻辑。

风险一：训练数据的合法性基础 (Art. 6) 正在坍塌

AI行业赖以生存的两大训练数据来源——"合法利益"和"公开数据"——的法律基础正被欧盟监管机构系统性地瓦解。这已成为AI产业面临的"棘手且基础性法律难题"。

• "合法利益" (Legitimate Interest) 的困境：

OpenAI声称依赖GDPR第6条第1款(f)项（合法利益）作为处理用户数据训练模型的依据。但Garante对此持否定态度，Garante在决定中表示，OpenAI在隐私政策中未对"合法利益"作出清晰明确的阐释（其表述过于简略），亦未赋予数据主体相应的反对权，这暗示了其在构建正当利益合法性基础过程中存在显著缺陷。欧洲数据保护委员会(EDPB)的工作组亦强调，这必须经过严格的利益平衡测试，充分考虑数据主体的"合理预期"。

• "公开可用数据" (Publicly Available Data) 的陷阱：

这是AI行业（特别是爬虫训练）的另一个常用借口。爱尔兰DPC对X(Grok)的调查正在正面挑战这一问题："公开可用"是否等于"可合法用于AI训练"？答案很可能是否定的。

风险二：算法"幻觉"被定义为数据"不准确" (Art. 5)

欧盟监管者正在将AI的技术性缺陷（幻觉,Hallucinations）法律化，将其定性为违反GDPR"准确性原则" (第5条第1款(d)项)。

Garante在OpenAI案中虽然将此问题的最终决定移交给了爱尔兰DPC，但其发表的看法极具指向性：

1.Garante承认模型存在固有缺陷，但强调ChatGPT提供的结果很可能被用户视为准确信息，无论其实际准确性如何。

2.Garante明确指出，OpenAI采取的"透明度"措施（如弹窗提示内容可能不准）有助于履行透明度原则，但不能豁免其遵守准确性原则的义务。

因此，Garante要求的是实际效果的准确性，而非仅仅是过程的透明。这对一个"概率性质导致模型会产生片面或有偏差" 的系统而言，是极难完成的任务，对AI企业的算法设计和数据治理提出了根本性挑战。

风险三：未成年人保护 (Art. 8)——重复出现的"高压线"

在所有风险中，未成年人保护是Garante执法的优先领域和重复出现的"高压线"。

Garante在封禁ChatGPT之前，就已于2023年2月禁止了AI聊天机器人Replika，主要原因就是其对未成年人构成重大风险且缺乏任何有效的年龄验证机制。

Garante对OpenAI案的论证逻辑尤其值得企业关注：

1.Garante援引GDPR第24条和第25条（设计和默认的数据保护），论证"实施差异化保护措施的先决条件是能够对不同人群予以有效识别"。

2.因此，进行未成年人年龄验证是必不可少的环节。

3.Garante进一步质疑，由于无法验证年龄，OpenAI无法确保与13-18岁未成年人的合同（服务条款）在意大利合同法下有效。因此，其以"履行合同所必需"（GDPR第6条第1款(b)项）作为处理未成年人数据的合法性基础亦存在瑕疵。

风险四：透明度义务 (Art. 12/13)——被低估的合规基础

OpenAI案表明，监管机构对于透明度的审查已进入"像素级"。Garante对OpenAI隐私政策的批评极为具体，揭示了AI企业在透明度上的常见疏忽：

- **语言**：截至2023年3月30日，仅提供英文版，对非英语用户（包括未成年人）不友好。
- **访问路径**：注册流程中链接位置不佳，用户无法在输入数据并创建账户之前阅读隐私政策。
- **内容**：

1.对非用户（训练数据来源）：对于用于模型训练的非用户的个人数据处理情况，未提供任何信息。

2.对用户（训练数据来源）：描述过于笼统（如"改进服务""开展研究"），用户通过描述无法知晓其个人数据将被用于算法训练。

Garante的关键逻辑驳斥了OpenAI称其通过技术文件、博客文章披露训练信息的抗辩：

1.Art.12/13的信息义务应是主动告知（Push），使相关方处于被动接收信息的地位；而非要求相关方主动探寻 (Pull)。

2.非用户既无动机、也无合理预期去查阅这些专业性极强的技术文件，以获知其公开数据被用于模型训练。

这为所有AI企业提供了明确指引：必须在单一、易得的隐私政策中，用清晰、具体的语言，向所有数据主体（包括非用户）说明其数据如何被用于训练。

005 > 结论：核心观点——合规必须前置于业务

从2023年对OpenAI的临时禁令，到2024年的1500万欧元罚单，再到2025年对DeepSeek的迅速封禁，欧盟对AI的监管执法已经完成了从"教育与谈判"到"常态化执法"再到"零容忍阻断"的演变。

AI 企业出海欧盟，如今面临两大清晰的致命风险：

1.**运营风险**：基于GDPR第66条（紧急程序）的"业务阻断"，可由单一 DPA 在数日内发起。

2.**地缘政治风险**：基于GDPR第44条（国际传输）的"数据主权"违规，监管机构对此"零容忍"，这对中国AI企业尤为致命。

Garante在OpenAI案中确立的"问责制"和"事前合规"逻辑，彻底终结了"业务先行，合规后补"的幻想。

因此，对于计划出海的AI企业，隐私合规（特别是GDPR）不再是法务部门的附加项或成本中心。它是一种前置于业务的、决定企业能否进入欧盟市场的核心战略前提。企业必须在R&D阶段就构建可证明的 (provable) 合规体系，否则出海企业将落入欧盟监管机构的雷达范围，可能随时被"锁定"而失去本应有的商业机会。

蔡鹏
合伙人
知识产权部
北京办公室
+86 10 5087 2786
caipeng@zhonglun.com

作 者 / 马远超

中国AI企业赴欧盟市场展业的IP合规风险与建议

001 > 欧盟《人工智能法案》带给中国AI企业赴欧盟展业的知识产权合规挑战

1.欧盟《人工智能法案》对提供者版权合规要求

《人工智能法案》第五章第2节第53条"通用人工智能模型提供者的义务"规定："通用人工智能模型提供者应：……制定尊重欧盟版权法的政策，特别是通过先进技术等手段，确定和尊重根据2019/790号指令（即《数字化单一市场版权指令》）第4条第3款表达的权利保留。"

《数字化单一市场版权指令》第4条规定："1.成员国应规定，以文本和数据挖掘为目的，对合法获取的作品或其他内容进行复制与提取的行为，属于96/9/EC指令第5条（a）项与第7条第1款，2001/29/EC指令第2条，2009/24/EC指令第4条第1款（a）和（b）项，以及本指令第15条第1款所规定的权利的例外。2.以进行文本和数据挖掘为目的，根据第1款复制和提取的作品或其他内容可保留到必要时为止。3.适用第1款规定的例外或限制的条件是，权利人没有以适当方式明确保留对上述作品或其他内容的使用，例如针对网上公开提供的内容采取机器可读的方式。4.本条不影响本指令第三条的适用。"《数字化单一市场版权指令》第4条主要涉及数字环境下版权的例外或限制，特别是在教学目的下的使用；旨在为教育机构在数字环境下使用作品提供了一定的灵活性，同时也保护了版权人的利益。这一条款有助于平衡版权人和教育机构之间的利益，促进数字环境下教育资源的合理利用和传播。由于AI模型的开发商往往是商业机构，用于商业用途，往往无法符合构成上述合理使用的前提条件。

欧盟版权法是一个复杂的法律体系。除上述《数字化单一市场版权指令》之外，还包括《版权保护期指令》《关于信息社会协调版权及相关权利某些方面的2001/29/EC号指令》（简称为《信息社会指令》），以及适用于欧盟的国际条约，例如《与贸易有关的知识产权协定》《伯尔尼公约》等。

欧盟《版权保护期指令》的主要内容涉及文学艺术作品的版权保护期限以及新闻出版者邻接权的保护期限。该指令主要内容有：

1）文学艺术作品的版权保护期：根据指令规定，文学艺术作品的

保护期为作者有生之年以及过世后70年。这一保护期不受作品合法公之于众的具体时间影响。这一规定是为了保护作者及其后来两代人的利益，考虑到共同体内人均寿命的增长，原有的保护期（作者有生之年加死后50年）已经不能覆盖两代人的时间跨度。

2）新闻出版者邻接权的保护期限：指令中明确规定了新闻出版者邻接权的保护期限。规定新闻出版者邻接权的保护期为新闻出版物出版后两年，从该出版物出版次年的1月1日起算。

此外，指令中还涉及版权例外情形、非流通作品的使用保护、延伸性集体管理制度、流媒体平台上的视听作品许可问题、公有领域的视觉艺术作品使用问题、在线内容分享平台的特殊责任机制以及作品、表演开发利用合同中对作者、表演者的保护等。

欧盟《信息社会指令》的主要内容涉及信息社会中版权和邻接权的协调，该指令主要内容有：

1）目的和范围：指令旨在协调信息社会中的版权和邻接权，以适应数字技术的发展和电子商务的需要；指令适用于在欧盟内部市场中提供的信息社会服务，特别是电子商务服务。

2）版权的保护：指令强调了版权的重要性，并要求成员国提供适当的法律保护，以防止未经授权的复制、发行、向公众传播等行为；指令还规定了技术保护措施的法律保护，以防止对版权作品的非法复制和传播。

3）权利管理信息的保护：指令要求成员国提供法律保护，以防止未经授权去除或改变电子权利管理信息，这些信息用于识别作品、作者、权利人或作品使用的条件和期限。

4）制裁和救济措施：指令要求成员国规定适当的制裁和救济措施，以应对侵犯版权和违反指令义务的行为；成员国应采取必要措施，保证权利人能够提起损害赔偿诉讼和/或申请禁令，以及在适当条件下申请没收侵权材料。

5）与其他法律条款的关系：指令明确指出，它不影响其他相关法律条款的适用，如专利权、商标、外观设计权等。

6）适用时限和过渡性规定：指令的条款适用于在指令生效后受成员国立法保护的作品和其他客体；指令的适用不影响指令生效前已完成的行为或获得的权利。

欧盟《人工智能法案》第五章第3节第55条"具有系统风险的通用人工智能模型提供者的义务"规定，具有系统风险的通用人工智能模型的提供者同样需要遵守上述第53条规定的义务。

值得强调的是，欧盟人工智能办公室将根据提供者公开提供的用于训练的内容摘要等，监督提供者是否履行了上述义务，而不对训练数据的版权合规性进行逐项核查或评估。

2.欧盟《人工智能法案》对提供者商标合规要求

《人工智能法案》第三章第3节第16条"高风险人工智能系统提供者的义务"规定："高风险人工智能系统的提供者应……(b)在高风险人工智能系统上标明其名称、登记商号或登记商标、联系地址，如无法标明，则在包装或随附文件上标明；……"

《人工智能法案》第三章第3节第23条"进口者的义务"规定："3.进口者应在高风险人工智能系统及其包装或随附文件，如适用，上注明其名称、登记商号或登记商标以及联系地址。"

《人工智能法案》第三章第3节第25条"人工智能价值链上的责任"规定："1.为本条例之目的，任何分销者、进口者、部署者或其他第三方均应视为高风险人工智能系统的提供者，在下列任何一种情况下，均应承担第16条规定的提供者义务：(a)在已投放市场或提供服务的高风险人工智能系统上冠以自己的名称或商标，但不妨碍合同中关于以其他方式分配义务的规定；"

根据《人工智能法案》上述规定，高风险人工智能系统提供者，包括任何分销者、进口者、部署者或其他第三方，都需要在高风险人工智能系统冠以自己的名称或商标。虽然对通用人工智能模型提供者没有提出明确的要求，但通用人工智能模型提供者通常都会标识自己的名称或商标。无论标识企业自己的名称、字号或者商标，都属于商标性使用，都存在商标合规风险。

商标不同于版权，具有地域性特征，即在一个法域内注册的商标仅在该法域内获得保护，当在另一个法域内使用该商标时，需要在另一个法域获得核准注册，否则不受法律保护。例如，中国的注册商标原则上不受欧盟法律保护，欧盟的注册商标原则上也不受中国法律保护，驰名商标除

外。为此，当提供者在欧盟境内提供人工智能模型服务时，应当在欧盟境内获得商标注册，至少不能与他人享有的在先欧盟注册商标相冲突。

002 > 中国AI企业进入欧盟市场的其他商标和专利风险排查

1.尽职调查欧盟市场的商标风险

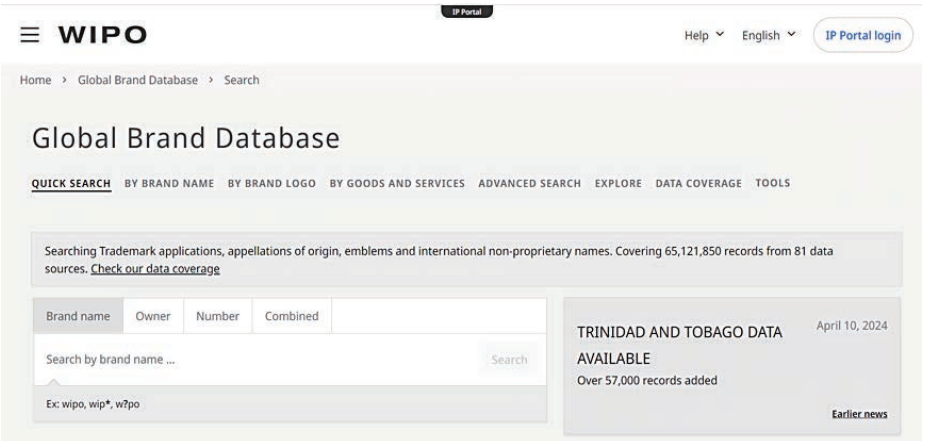
中国人工智能企业的人工智能产品在进入欧盟市场之前，需要做好商标风险的尽职调查。

调查的对象包括：企业的字号、商标、产品名称、软件的LOGO、软件名称、广告语。

企业既可以委托专业的法律服务机构完成尽职调查，也可以自行到官方或者商业商标数据库进行检索。

常用的官方商标数据库包括：

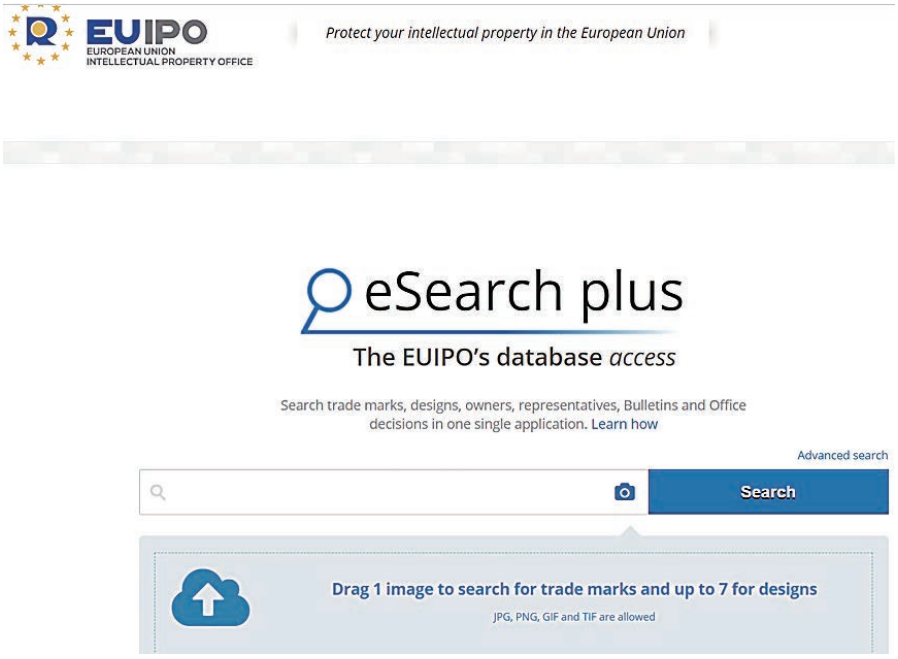
世界知识产权组织（WIPO）提供的Global Brand检索数据库，网址：https://www3.wipo.int/branddb/en/。Global Brand检索数据库收录有全球58个国家或地区的商标数据，WIPO作为全球性的知识产权国际组织，其建立的有特色的商标数据包括：WIPO Emblems (Article 6ter)商标、WIPO Appellations of Origin (Lisbon) 商标、WIPO International Trademarks (Madrid)商标；其中FILTER By中的Image标签可以导入图片进行过滤检索。



欧盟知识产权局（EUIPO）提供的TMview检索数据库，网址：https://www.tmdn.org/tmview/#/tmview。TMview检索数据库共收

录有69个国家或地区的商标数据，比Global Brand检索数据库覆盖的国家或地区多11个。TMview检索数据库支持中文检索界面，界面友好，Trade mark type字段中可以对立体商标、颜色商标、颜色组合商标等进行精细检索，也支持用户通过上传4M的静态或者动态图片/照片（支持jpg、png、gif、tiff格式）的方式进行检索。

欧盟知识产权局（EUIPO）提供的eSearch plus检索数据库，网址：https://euipo.europa.eu/eSearch/。eSearch plus数据库提供一站式的欧盟商标、外观设计、权利人、代理人、官方公报（商标、外观设计官方公报）、官方决定的整合检索，同样支持用户通过图片/照片的方式对商标/外观设计进行图形检索。



如果企业发现自己的字号、商标、产品名称、软件名称、LOGO、广告语等构成商标性使用的要素，与第三方已经申请或者已经获得授权的商标在相同类别商品（包括服务）或者类似商品上相同或者近似，就应当引起重视，在权衡利弊后，需要决定是否调整自己在欧盟的商标标识。

2.在欧盟申请注册商标

欧盟的商标注册实行"先申请先保护"原则以及"双重保护"原则。"先申请先保护"原则是指，在先申请人可以反对他人在后对相同或近似商标的混淆性使用或者注册申请。"双重保护"原则是指，欧盟注册商标与欧盟成员国注册商标可以共存，两者受保护力度完全相同，两者相互构成抵触他人申请注册商标的在先权利。企业直接向欧盟知识产权局提交商标注册申请，与在欧盟成员国分别提出商标注册申请相比，更具优势，因为费用更低，可以节省翻译和管理费用。欧盟注册商标有效期为10年，可提前6个月提交续展申请，续展宽展期为6个月。欧盟知识产权局提供的所有商标服务具体内容和费用都可以在官网查询，网址：https://euipo.europa.eu/ohimportal/fees-payable-direct-to-euipo。

如果以中国企业名义申请欧盟商标，根据欧盟知识产权局的要求，必须指定代理人，代理人可以是：

1）欧洲经济区内有权限的商标代理机构/律师事务所（Legal practitioner）；

2）被欧盟知识产权局许可的专业代表（Professional representative）；

3）在欧盟有住所/主要营业地/工商业机构的自然人或法人的员工（Employee of a natural or legal person）。

如果以注册在欧盟的企业名义申请注册欧盟商标，无须指定代理人就可以直接申请，并视申请人的性质（自然人或法人）提供对应的基本信息即可。

3.在欧盟的专利风险排查

根据国家工信安全中心、工信部电子知识产权中心2024年4月10日发布《新一代人工智能专利技术分析报告》，当前，我国人工智能创新链不断完善，AI基础层、模型层创新活跃，呈现多线发展，生态联动的创新态势。截至2023年底，新一代AI基础层（AI芯片、AI框架）、模

型层中国公开专利达到6.2万件，其中有效专利近2万件，审中3.5万件。2017年以来，专利申请年均增长率超过43%，自然语言处理和计算机视觉占比最高，也是目前发展最为迅速，产业应用较快的技术分支。我国申请人新一代AI专利在华申请数量已形成主导优势，百度、华为、腾讯和阿里巴巴等能有效汇聚高质量数据、大规模算力和先进算法的新一代AI全栈型企业正成为AI创新布局的"主力军"。

在欧盟，IBM、谷歌、英特尔、三星电子、微软公司、松下电器等国际科技巨头也早已申请了大量AI技术领域的相关专利。相关技术主题涵盖语音识别、大数据和云计算、机器学习、自然语言处理、计算机视觉、具体应用六个方面。这些国际科技巨头已经在AI专利进行了全产业链的多方位布局。

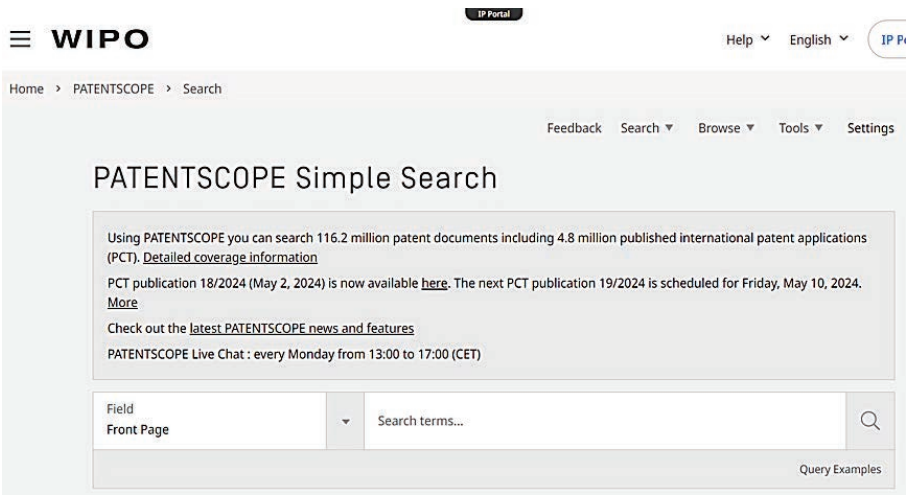
为此，中国人工智能企业的人工智能服务一旦进入欧盟市场，专利合规风险是不可回避的问题。

对于中国人工智能企业而言，在欧盟进行专利风险排查，是一项非常复杂的法律工作。企业应当委托中国以及欧盟专业的知识产权律师，在企业的技术人员或者知识产权专员的配合下，共同完成这项工作。

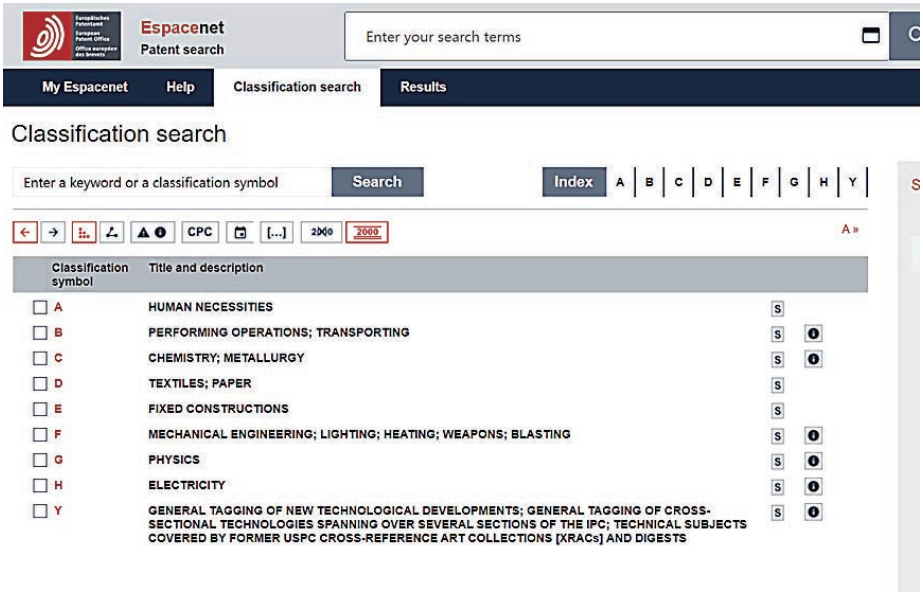
企业应当结合人工智能产品使用的技术方案、方法、外观设计、用户界面等进行全面检索、分析。

常用的官方专利数据库包括：

世界知识产权组织（WIPO）世界专利数据库，网址：https://patentscope.wipo.int/search/en/search.jsf。用户可以对不同国家或者地区专利局的专利文献进行查询。



欧洲专利数据库，网址：https://worldwide.espacenet.com/patent/。用户可以检索在欧盟地区申请的专利，查看法律状态等信息。



003 > 中国AI企业赴欧盟市场展业的合规建议

1.组建海外知识产权管理内部人才以及外部供应商体系

知识产权管理具有较高的专业门槛，涉及版权、商标、专利、商业秘密等多项部门法。海外知识产权管理不仅涉及语言门槛，还涉及不同法域、不同法系的专业门槛。为此，企业需要培养、组建专业的懂法律、懂外语、懂管理的知识产权管理团队。

在处理海外知识产权监测、申请、布局、分析、纠纷应对等事项时，不可避免地需要国内外专业律师的支持和协助。在国内外挑选专业过硬、反应及时、收费合理的律师事务所，并非短期内可以实现，需要通过不断磨合、试错才能找到适合自己企业需求的律师事务所合作。

2. 提前进行知识产权海外布局

人工智能产品研发与知识产权海外布局，应当同步进行。中国企业需要根据产品海外规划、市场拓展计划，前瞻性地进行欧盟、美国、日

本、韩国等主要经济体的知识产权海外布局，具体包括商标和专利的申请注册、海外知识产权法律环境的研究、海外竞争对手的知识产权布局监测等。

3. 加强知识产权风险预警监控

就专利风险而言，在人工智能产品研发阶段、选择产品上市地域之前，中国企业应当通过专利检索掌握各法域内专利布局状况，识别和规避潜在专利风险。企业应当聘请专业机构或专家就特定的他人专利进行自由实施分析（FTO），分析判断是否存在实质性专利侵权风险。

就商标风险而言，企业对人工智能产品服务地域范围内的商标注册情况需要进行动态监测。一方面关注自身的商标性使用标识是否已经在当地核准注册、以及做好事后权利维护工作，另一方面关注他人是否存在申请注册类似或者近似商标的行为，及时通过撤销、异议、无效等程序阻止他人的恶意申请。通过对他人申请注册商标的监测，也是发现仿冒产品的途径之一。

就版权风险而言，企业需要事先研究训练数据来源地的版权保护法律环境。在欧盟，《数字化单一市场版权指令》对人工智能获取、使用他人数据进行了严格限定，相比美国而言监管严格。中国企业应当持续关注欧盟范围内的人工智能语料收集、使用的司法审判、行政执法案例。

4. 增强海外知识产权纠纷应对能力

一方面，中国企业遭遇知识产权侵权行政执法（临时禁令扣押、没收、搜查等）时，应沉着冷静，首先在现场应当依法配合执法人员的调查，现场及时寻求当地专业律师的法律帮助。事后，中国企业可在当地专业律师帮助下依照当地法律向相关部门提出异议、申诉等，并积极收集、提交未侵权或者情节从轻的相关证据。

另一方面，中国企业如果因知识产权侵权遭遇海外诉讼狙击，首先应当及时咨询中国境内律师，对诉讼进行初步分析判断，其次在中国境内律师的帮助下、积极挑选境外应诉律师团队，对诉讼进行进一步分析判断。与此同时，企业组建由企业法务人员、中国境内律师、境外律师团队共同组成的应诉团队。通过对诉讼发生地的法律环境规则的了解，

对诉讼程序周期的了解，对证据的收集准备，对诉讼后果的预判，制定筹划详细的诉讼策略。最后，中国企业也可以选择协商谈判的恰当时机，还可以寻求中国商务部或者其他政府部门、行业协会等组织的帮助和支持。

马远超
合伙人
知识产权部
上海办公室
+86 21 6061 3188
myc@zhonglun.com

作 者 / 龚乐凡 高翊菲

AI深度伪造——人工智能引爆欺诈危机与境外追索的破局之路

> "旧世界分崩离析，新时代正在光速到来。"
>
> ——梁文锋

根据德勤的一项统计，高达25.9%受访企业高管提及他们在过去的2024年里经历过一次或者多次深度伪造（Deepfake）事件。[1]德勤的研究部门预测，到2027年生成式AI在美国导致的欺诈损失将达到400亿美金，这与2023年相比123亿美金相比，足足增长了约2.25倍。

AI技术正在重塑现实与信任边界，从香港奥雅纳公司遭遇AI换脸高管骗走2亿港元，到诈骗集团利用伪造"美颜"人设跨境收割虚拟货币投资者；从"AI马斯克"以假乱真掏空美国老人毕生积蓄，到恶意软件"GoldPickaxe"突破生物识别防线窃取财富，再到南非冒用商界领袖形象的深度伪造投资骗局——这些案例警示着AI时代的人类：人工智能已不仅仅是工具，同时也逐渐演变为欺诈犯罪的颠覆性变量，其以假乱真的"公信力嫁接"与跨境匿名特性，有将全球金融安全推向重构边缘的风险。尽管如此，在法律与风控视角下，我们仍可将其视为一种技术滥用行为，并通过系统性手段进行遏制与追索。

本文立足于这场由代码与人性交织的新型战争，剖析Deepfake诈骗的运作模式，并探讨如何构建"技术防御-法律规制-国际协同"的三维防线。同时，本文还聚焦跨境追索这一关键战场，旨在为潜在受害者提供一套融合数字取证、国际司法协作与资产追踪的果断反击路径，帮助其在虚拟"迷雾"中获得正义。

1.See *How AI is Redefining Fraud Prevention in 2025*，https://www.threatmark.com/how-ai-is-redefining-fraud-prevention-in-2025/.

> **“生存下来的不是最强壮的物种，也不是最聪明的，而是最能适应变化的。”**
>
> ——达尔文

境外追索的一大实战场景，来自于近年来深度伪造（Deepfake）技术通过高度逼真的音视频合成能力，模仿人脸人声以进行金融诈骗犯罪。此类技术不仅突破了传统身份验证的防线，还因其跨国性、隐蔽性等特点，使得追查和追索面临前所未有的挑战。

（一）假面幻影：AI换脸的惊天骗局[2]

2024年1月中旬，奥雅纳香港分公司收到来自英国总部的首席财务官的紧急通知，要求进行一项高度机密的交易。为确保交易的顺利进行，香港团队迅速召集了多名财务部职员，通过视频会议的方式共同商讨。

实际上，犯罪嫌疑人通过公司的YouTube视频和从其他公开渠道获取的媒体资料，利用Deepfake技术成功地仿造了英国公司CFO的实时视频和声音，还伪造了多名高层管理人员以及同事的形象和声音，假冒跨国企业奥雅纳（Arup）英国总部的CFO及高管团队，参与了一场由Deepfake技术伪造的多人视频会议。会议中，诈骗团伙利用公开资料合成高管的面部表情和声音，要求员工向指定账户转账。员工在未核实身份的情况下分15次总计转出2亿港元，一周后该员工与总部联系时，才发现被卷入骗局。这一情形并非传统意义上的网络攻击，因为系统并没有遭受黑客等入侵，犯罪嫌疑人利用Deepfake技术来制造紧迫感，规避企业内部沟通的安全屏障。该跨国公司的首席信息官表示，跨国企

2.衢州智造新城公安：《震惊！“变脸”冒充CFO，骗走两个亿！香港最大AI诈骗案细节曝光》，https://mp.weixin.qq.com/s/oZefx4UsmiKeSZOyP8-xNA.

业经常遭受包括Deepfake在内的网络欺诈，近年来网络欺诈的数量以及复杂程度都急剧上升。

该案暴露了跨国企业的两大合规风险痛点，一是企业对远程身份核验以及可疑交易的疏忽，二是跨国企业内部沟通机制存在断层。诈骗者利用时差和信息差，在短时间内完成资金转移，而受害者往往缺乏对于该等新型诈骗的认知，并因"紧急指令"的压迫感而放松警惕，最终遭受严重经济损失。

（二）"美颜"陷阱：中国香港"虚拟货币投资"诈骗案[3]

2025年1月6日，中国香港警方捣破以Deepfake技术在社交平台诱骗他人投资虚拟货币的诈骗集团，涉案金额约3400万港元。

该案中，诈骗集团招揽想"赚快钱"的年轻人加入，训练他们以虚假人设在交友平台开设账户，假装外貌出众、生活奢华，以此结识我国台湾地区及东亚境外地区的人士，按剧本聊天。在了解对方背景后，再利用Deepfake技术进行视频通话获取信任，进而声称虚拟货币有可观回报，诱导受害者在虚假的虚拟货币平台投资。诈骗集团在收到资金后则会立刻转走，并与受害人断绝来往。

（三）"AI马斯克"投资骗局[4]

美国某官方数据显示，2023年，美国60岁以上老人因各类欺诈骗局造成的损失达到283亿美元（2053亿人民币），其中一部分骗局由AI协助完成。[5]在这些AI协助完成的案件中，有不少是利用名人的公信力，定向投放深度伪造视频至易受骗群体进行诈骗，如"AI马斯克"投资骗局案。

2024年8月，埃隆•马斯克因Deepfake卷入了一场诈骗事件。当时82岁的退休老人史蒂夫•比彻姆正在观看一段视频广告，在广告中，"马斯克"正在推销一项承诺具有可靠快速回报的投资项目。由于广告中，

3.南方都市报，《香港一诈骗集团以交友诱骗投资，被捣破！涉案人员含港超球员》，https://baijiahao.baidu.com/s?id=1820496292508438787&wfr=spider&for=pc.
4.每日经济新闻，《高达近500万元！"假马斯克"骗光82岁老人毕生积蓄，口型一样，甚至还有南非口音》，https://mp.weixin.qq.com/s/h7VBG7NZmrFeLQ4-lqbL9g.
5.https://press.aarp.org/2023-06-15-AARP-Report-Finds-28-Billion-a-Year-is-Stolen-from-US-Adults-Over-60.

"马斯克"面目形象与南非口音极其真实，该老人对广告信息深信不疑，于是便联系了"马斯克"所背书的这家名为Magna-FX的外汇公司，开设了一个248美元的账户。在此之后，该老人被频繁诱导着进行一系列的转账交易，直至最终耗尽了自己退休金账户的69万美元。

据悉，该诈骗者是基于马斯克的一段真实采访视频，利用Deepfake技术对其嘴部动作进行了唇部同步技术编辑，并附加AI加工后的声音，使其与他们为数字人编写的虚假剧本完美匹配。此案揭示了Deepfake技术的"公信力嫁接"特性——利用名人效应降低受害者戒备心，并通过小额试水逐步扩大诈骗金额。

（四）"人脸劫持"：人脸识别系统的突破[6]

2023年10月，国外网络安全公司Group-IB发现了一个能够窃取、收集人脸识别数据的银行木马程序"GoldPickaxe"。该恶意软件的iOS版本，诱骗用户进行人脸识别、提交身份证件，从而窃取个人生物信息如面部识别数据、银行账号、电话等身份信息。随后这些敏感信息会被转换成人工智能生成的深度伪造图像，其能够成功突破人脸识别系统，实施诸如登录用户的银行账号进行转账等下游犯罪操作。

"GoldPickaxe"也有安卓版本，功能更多。除收集盗取受害人的人脸信息、身份证、银行卡等信息外，安卓版还能在手机的相册中检索人脸图片，并向银行App索求人脸识别授权等服务。

更具隐蔽性的是，此类犯罪行为常见的攻击策略是组合使用短信轰炸和网络钓鱼技巧，并经常伪装成政府服务代理（如数字养老金和政府信息门户网站），受害者往往难以识别该类软件，从而蒙受损失。

（五）南非投资骗局[7]

南非金融部门行为管理局（FSCA）警告公众，不要接受Gold Earnings Hub和Africa Gold Capital提供的财务建议、帮助或投资机会。这两

6.GoUpSec,《人脸识别要完？首个"人脸劫持"银行木马诞生》，https://mp.weixin.qq.com/s/7pxgJfiPo3HONc3exoXEjw.
7.《细思极恐！深度伪造Deepfake 投资骗局蔓延南非甚至针对南非亿万富翁！》，https://mp.weixin.qq.com/s/XSXBVkY_58reHBNn7bjUKQ.

家公司被发现使用Deepfake技术获取投资。根据FSCA披露的信息，这两家公司未经授权提供金融服务，并利用Deepfake图片和视频冒充知名人物，如非洲彩虹矿业（ARM）执行主席帕特里斯•莫特赛佩（Patrice Motsepe），并利用伪造形象宣传两家公司的投资项目。

002 > 深度伪造与跨境资产追索：四大核心迷思与误区

（一）迷思误区之一：以为资金"人间蒸发"——认知误区的局限

实施境外追索最大的挑战和障碍，是当事人的认知和判断误区。针对资金或者加密货币被骗被盗，当事人一个最大的误区是，认为资金早已转走不知所踪，想追回就是"大海捞针"，并且需要"国际刑警"加本地警方的联合行动，更是难上加难。因此放弃采取行动，从而错过了黄金窗口。

其实，资金从未"蒸发"，只是你没有发现而已。由于资金是通过转账的方式被转移的，大概率会留下资金转移的轨迹。以Deepfake方式进行欺诈，本质上与其他网络黑客欺诈并无区别，只是AI技术手段更加高超，更难被识破。传统的典型的BEC （business email compromise 即商业电邮侵入方式）的网络黑客欺诈模式，是通过侵入受害人公司或者个人的电脑系统，冒充相关的"收款方"当事人发送电邮、指令，要求将汇款账户进行调整，最后导致资金汇出后"石沉大海"，而实际"收款方"根本没有收到款项，与"付款方"核对之后才发现自己被骗。

换句话说，这种欺诈模式早已有之，想要成功实现跨境追索，就得克服认知障碍。许多企业在遭遇这类诈骗时，由于实施欺诈者"隐匿遁形"，受害者在报警之后，被告知欺诈行为发生地和诈骗者都在境外，难以受中国法律的管辖和保护，就想当然地认为在境外追回款项几乎是"不可能完成的任务"。

实际上，无论是资金还是比特币，如果找对专业机构，包括境内的法律顾问和境外的专业机构，采用正确的追索方法，追索甚至找回被转移到境外的资产是有可能实现的，甚至还可能获得额外的补偿和赔偿，例如利息损失、包括律师费在内的追索成本等。

（二）迷思误区之二：六神无主，错过黄金24小时

与线下的盗窃抢劫活动不同，网络电信诈骗、AI深度伪造的诈骗作案人经常隐匿无踪，受害人对技术不了解、高估其难度，因此在事件发生之后六神无主，病急乱投医，也错过了最佳的追击时机。

诈骗资金往往通过加密货币或离岸账户多层洗白。例如，“假面人会议”案中，2亿港元在1周内被转移至多个境外空壳公司账户。**传统司法互助请求（MLA）流程冗长，难以满足快速冻结资产的需求。**在资产追索领域，有一个“**黄金24小时**”的说法，意思是在欺诈行为被发现后，企业需在24小时内采取紧急措施，以遏制损失的扩大。

这一时期的关键性在于，欺诈资金通常在短时间内被快速转移，最终流向难以追踪的海外账户或加密货币地址。通过迅速冻结涉及的账户、联系支付网关并采集证据，企业可以有效阻断欺诈资金的转移路径，最大程度降低损失。企业如何把握这个“黄金24小时”，不仅是损失控制的关键，也是提升企业应急能力的重要环节。通过快速响应，企业可以显著降低直接损失的金额，同时积累应对类似事件的经验和能力。

（三）迷思误区之三：高估诈骗行为处置的复杂性，认为法律手段难以解决

Deepfake诈骗常涉及多国主体，例如诈骗者在A国伪造身份、服务器位于B国、资金流向C国。各国法律对“虚拟身份犯罪”的定性不一，且此类行为往往会落入刑民交叉的领域，刑事追责与民事索赔程序存在重叠。受害者往往非常容易陷入迷茫，不知道先选择民事诉讼还是诉诸刑事追索手段，也不知该先开展境内追索还是境外追索。一旦**选错或盲选追索方案，极易导致各国司法机关的管辖权争议，造成不必要的时间窗口的错失以及最终的资金损失。**

所以，如果能提前进行布局、制定预案并演练，在事发之后迅速采取行动，可以为后续的法律诉讼和追讨损失提供坚实的基础。

而实践中，“反常识”的地方在于，在海外，在不同法域非常“管用”的司法救济工具在中国并不适用，导致大多数国内的司法人员以及律师也会误认为，应对跨境的电诈并追回损失会非常困难。实际上，相较于可能花费五六个月才可能决定是否立案并投入精力的海外警察，境内外

专业律师的快速联动往往能够产生更好的追索效果。

然而不少境外的律所（包括一些非常知名的国际大所），并不了解资产追索这一"小众"而又"高精尖"的业务领域。**这就非常仰赖在这方面有专业能力的法律顾问**，能在最短的时间内快速反应，找到最合适的境外法律服务机构和调查机构，主导整个境外追索的流程，启动被骗资金的境外追踪。

（四）迷思误区之四：面对举证责任"望而生畏"

客观而言，遭遇Deepfake以及其他类似的网络欺诈，证据固定与鉴定的复杂性确实存在。Deepfake视频的鉴定需依赖专业技术团队，而跨境案件中，证据的合法性常因各国电子证据标准不同而受质疑。例如，"AI马斯克"案的侦破需协调美国、南非等多国技术机构协作鉴定视频真伪。因此，在跨境追索领域，如何保存所有相关数字证据，包括视频、音频、聊天记录和系统日志等，以确保其原始格式和完整性？如何保留涉案物理设备，避免操作以防止数据破坏？如何进行证据的公证认证？如何详细记录所有操作步骤，形成完整的证据链？这些问题，都需要境内外的跨境法律专家、资产追索专家的协助。

然而，对此望洋兴叹无济于事。事实上，普通法系的国家和地区，包括常见的离岸小岛BVI、开曼、百慕大，在证据制度上为被欺诈的受害人提供了诸多便利，例如向法院申请出具调查令、冻结令以及证据开示制度（要求对方提供证据）等。这些工具让海外的资产追索在很大程度上，并不那么依赖警方支持与大量证据的搜集，这与国内的司法和证据制度相比，可以说是"反常识"的。

003 > 鹰扬虎视：从深度伪造骗局中解锁防范密钥

> **"像专家一样掌握规则，才能像艺术家一样打破规则。"**
>
> ——毕加索

在AI时代浪潮席卷之下，深度伪造技术犹如一把双刃剑，在带来创

新与便利的同时，也衍生出日益严峻的风险隐患，深度伪造风险已然成为企业和金融机构在网络空间中必须直面的重大挑战。

针对AI时代所带来的深度伪造风险，从风控和法律角度，究竟该如何应对？笔者结合已有欺诈案例的特点，拆解企业内控的最佳实践，以及欺诈事件发生后的跨境追索法律实践，对于企业和金融机构防范和应对网络欺诈，无疑具有极为重要的参考价值和实际指导意义。

（一）乘虚而入——Deepfake攻击人性三大弱点

笔者通过分析已有的欺诈案例，梳理深度伪造欺诈的典型特征与作案手法，发现攻击者往往抓住企业员工的三个永恒弱点：层级压力、保密指示和制造紧迫感。

1. 层级压力（Hierarchy）

越是等级分明的企业，越容易中招。比如下级绝对服从上级的公司（许多东亚文化圈企业）中常见的情景："我是老板，我现在要你转一笔紧急款项。""不要问别人。""你要对我负责。"一旦加上"视频会议+声音克隆"，再谨慎的人也会松懈。

2. 绝对保密（Confidentiality）

诈骗常出现这样的指令："这是一个机密项目。""不能和 CFO 说。""只有你和CEO 知情。"这种"被信任感"反而让受害者更容易掉入陷阱。

3. 制造紧迫感（Urgency）

诈骗永远强调速度："必须20分钟内处理！""错过这次，收购案就完了！""银行关门前必须完成！"

一个看似真实到无法辨别的"CEO+绝对保密+紧急转账"，一旦这三者叠加，无论是跨国企业、家族办公室还是中小型企业，都可能中招。所以，基于上述特点，再借鉴企业内控领域的最佳实践，可以为企业构建坚固的风险防线提供宝贵的思路与方法（见下文）。

（二）铜墙铁壁——防范Deepfake欺诈的四道防线

为反欺诈建立铜墙铁壁，企业与个人需要建立应对Deepfake的四道防线——结合欺诈案例的分析与笔者的反欺诈实践，笔者总结出对企业客户、家办与高管团队最实用的一套体系。

1. 第一层：制度防线（Governance）

第一道防线，就是要建立制度和铁律——任何时间、任何场景，禁止"单人指令汇款"，这包括：

一方面，形成一条绝对铁律——无论视频会议中看到的人看起来多像CEO，都绝不能在未经二次验证的情况下转账；另一方面，强制要求员工必须通过第二通讯渠道核实（如拨打本人常用手机），核验之后方能转款。

也就是说，所有付款都必须经双人审批（Two-Man Rule）。

如果供应商更改账户，必须通过独立渠道验证，关键指令不可通过语音、视频直接执行——这条制度比任何技术都更有效。

2. 第二层：人员防线（Human Risk）

欺诈分子利用人性弱点，通过权威（无论是CEO还是公检法）的强势来施压——所以，公司需要打造一种"可以怀疑老板"的文化。

反欺诈专家特别强调，公司必须建立"无责上报"的文化。没有无责上报文化，就没有安全。为什么？因为 Deepfake 欺诈的核心之一就是让你"不敢怀疑"。

这点可能对大部分东亚企业尤其困难。很多员工害怕说："老板，我要确认一下。"因此企业必须让员工知道："怀疑可疑指令"不会被责难和报复。

核实流程是为了保护公司，任何高管必须接受被复核，只有这样，才能让员工在危险瞬间踩下刹车。

3. 第三层：流程防线（Process）

企业需深刻认识到，仅靠纸面制度远远不够，必须提前开展场景模拟演练。专家明确指出，一支经过演练的团队，其应对效能远超一份被束之高阁的政策——前者的实战价值可能是后者的百倍。正因如此，企业有必要将"深度伪造诈骗应急演练"纳入年度必修课，至少每年开展一次，切实筑牢防诈"实战防线"。

演练内容建议包括：

- 若 CEO 的视频会议里要求紧急付款，财务人员该怎么办？
- 若家族办公室收到"家族成员本人语音"要求转账该怎么办？
- 是否知道如何第一时间冻结资金？
- 是否知道如何保留会议视频作为电子证据？
- 是否知道如何联系银行进行SWIFT冻结？

如果不进行演练，危机来临时根本来不及反应，不知道哪些步骤可能出现卡点，从而错过"黄金24小时"的追偿时段（见下文）。

4. 第四层：技术防线（Technology）

针对Deepfake这样的AI技术，建议部署Deepfake实时检测工具以及其他软件进行防护，建立一道技术防线。

技术专家指出，人类肉眼99%无法识别 Deepfake。唯一有效的方法是技术检测。现有成熟技术能在Zoom、Teams等会议中嵌入，从而：

- 实时检测脸部是否为深度伪造；
- 识别语音是否为克隆；
- 自动发出警告；
- 自动记录可疑画面供取证。

未来，这样的工具会像"杀毒软件"一样成为企业的标配。

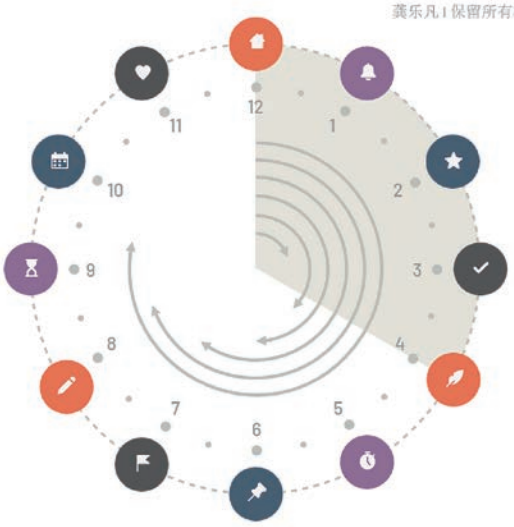
（三）欺诈应对——黄金 24 小时追偿行动指南

> **"世界上90%的问题都是有答案的，甚至是有标准答案的。"**
>
> ——雷军

遭遇Deepfake欺诈被骗之后，首先应该避免六神无主和手忙脚乱。正如雷军所说，"世界上90%的问题都是有答案的，甚至是有标准答案的。"Deepfake欺诈并非"无解"，关键是看认知和速度。

因此，在遭遇欺诈、资金转出时，可以参考执行黄金24小时追偿行动指南（企业与个人建议收藏）。

资金在银行体系中移动的速度有限，前24小时是追回资金的关键窗口。需要采取的步骤如下：



第1小时：冻结资金（最关键的一步）

立即通知本行（发送方银行），让银行通过SWIFT发出冻结请求，同时通知收款银行（Receiving Bank），要求立即冻结收款账户的所有资金。

第1-6小时：启动法律与技术团队

保存所有视频、音频、聊天记录等信息，快速联系熟悉该领域的专业律师，准备快速申请法院紧急冻结令、申请第三方披露（追查骗子账户开户资料）。

第6-24小时：开始跨境资金链追踪

通过专业律师快速对接不同法域的资产追索的团队，通过各种手段冻结资金账户，欺诈资金first landing account（首落账户）最可冻结，速度决定能否追回。

遭遇Deepfake：黄金24小时的追偿行动

1. 第1小时：冻结资金（最关键的一步）

立即通知自己的开户本行（发送方银行），让银行通过SWIFT（环球同业银行金融电讯协会）发出冻结请求，同时通知收款银行（Receiving Bank），要求立即冻结受款账户。速度越快，追回成功率越高。之后，第一时间联系专业的跨境追索法律团队。

2. 第1-6小时：启动法律与技术团队

采取进一步的行动，包括：保存所有视频、截图、音频、聊天记录，抓取所有网络日志与设备信息，联系资产追索律师，申请法院紧急冻结令（紧急情况下，可在法官的非工作时间与之取得联系），申请第三方披露（追查骗子账户开户资料），等等。这一系列的组合拳，需要和境外的相关司法领域的律师快速联动，与时间赛跑。

3. 第6-24小时：开启跨境资金链追踪

通常而言，因欺诈而汇出的资金的"首落账户"（first landing account）最容易冻结，之后会被转入一个到多个"中转站"账户，最终可能进入加密货币或地下钱庄，所以，速度决定能否追回。这往往依赖多个法域的专业"资产追索"律师的配合以及案涉法院和银行的配合。既然找到"首落账户"很关键，同理，找到能够链接境外律师采取联合行动的"首席"法律顾问同样关键。

在黄金24小时之后，不见得无法追回，但是难度和成本会变得更高。

004 > 小结：未来已来——AI对法律风控的颠覆性影响

> **"生活的本质是管理风险，而不是试图消灭风险。"**
>
> ——Walter Wriston（美国知名银行家，花旗集团前董事长兼CEO）

写给企业家的最后一句话——Deepfake欺诈带来的本质威胁不是技术，而是"我们再也无法相信我们看到的、听到的、甚至正在开会的人。"在这个时代，企业和个人需要的，是一套新的"信任系统"和风控系统。

人不能完全相信，视频不能完全相信，声音也不能完全相信。我们必须通过"制度+技术+流程+人"的企业风险防控体系才能共同建立新的"可验证信任"。这将成为未来5-10年众多企业、家族办公室和投资机构的核心议题。

根据笔者与从事跨境追索业务的国际律所的合作与交流的经验，无论是资金还是比特币被盗、深度伪造还是遭遇黑客侵入，又或是其他电信诈骗，想要追回被骗的资金，首先要破的是"认知的局限"——其并非"难以上青天"，而是完全可以通过找对专家、找对方法、找对路径，让追踪、查获和冻结被骗资产成为可能，挽回相关损失。

Deepfake技术的滥用不仅是法律问题，更是对跨境交易信任体系的挑战。唯有通过技术升级、专业的法律服务与跨境协作的三位一体，才能构建起有效的反诈防线、追回损失的资产。对于已遭受损失的受害者，及

时寻求专业跨境法律团队的帮助，与时间赛跑，抢在犯罪嫌疑人的前面，拦截和冻结资金，是挽回资产的关键一步。

（段慧明、罗凯文对本文亦有贡献）

龚乐凡
合伙人
私募基金与资管部
上海办公室
+86 21 6061 3608
lefangong@zhonglun.com

chapter 05

AI赋能律师行业

ai empowers the legal profession

作 者 / 韩梅

律师事务所部署AI之实践路径初探——从保密义务角度出发的探讨

001> AI时代的保密义务

保密义务是律师的核心义务之一，也是建立律师与客户之间信任关系的基石。《律师法》第38条规定："律师应当保守在执业活动中知悉的国家秘密、商业秘密，不得泄露当事人的隐私。律师对在执业活动中知悉的委托人和其他人不愿泄露的有关情况和信息，应当予以保密。……"律师的保密义务不仅源于法律规定，更根植于律师作为受托人的本质属性。

随着基于Transformer架构的大语言模型所代表的生成式人工智能（Generative Artificial Intelligence）的兴起和爆发，保密这一律师的传统义务面临着新的挑战。无论是使用DeepSeek等通用大语言模型服务或是基于大语言模型的垂类法律AI 产品，律师均需要输入有关信息。此类人工智能服务需要依赖云端的计算能力和数据处理基础设施，这意味着输入的信息可能需要离开律师事务所的本地环境，在云端进行处理和暂存。

这种技术架构上的特点，是否意味着律师对于人工智能服务的使用，将涉及保密义务的潜在违反？

笔者认为，保密义务的核心要求，并非是客户信息不得离开律师事务所这个"域"，而是律师事务所是否采取了合理措施防止信息的不当披露或使用。

在电子邮件系统早已成为通用办公工具的当下，业界从不会单纯因为律师事务所使用微软Microsoft 365的云端邮件系统认为律师事务所违反了保密义务。在具备合理技术与合同保障的条件下，这一逻辑可类推适用于人工智能服务。人工智能服务亦是律师事务所在提供法律服务过程中借助的技术工具，笔者认为，这一问题的关键应是如何在法律框架下安全地使用人工智能技术这一风险管理问题。律师事务所需要在安全保障、技术可行性和成本效益之间寻找平衡点。

那么，在保密义务这一红线之下，律师事务所究竟应当如何选择和使用人工智能服务？美国律师行业在保密义务与技术利用二者关系这一问题上的探索和演进，或许可以为我们提供一些启发。

当前，美国在人工智能技术和法律服务市场方面仍处于领先地位。笔者将在本部分梳理和探讨美国律师协会（American Bar Association，即"ABA"）关于律师保密义务的职业规则，以及部分知名国际律师事务所对于人工智能服务的采用情况。

1. 保密义务

ABA《职业行为示范规则》（Model Rules of Professional Conduct）[1]第1.6条（Confidentiality of Information）（a）款规定，除非客户做出知情同意或存在特定情形，律师不得披露与客户代理相关的信息[2]。这一义务与中国律师所承担的保密责任并无本质差异。

2. "合理措施"的引入与解释

2012年，ABA对《职业行为示范规则》进行了修订，在第1.6条中增加了（c）款，明确要求律师应采取合理措施防止与客户代理相关信息的无意或未经授权的披露或访问[3]。这一修订的关键词是"合理措施（reasonable efforts）"，其实际并未改变律师的保密义务，而是提醒律师对于技术带来的风险和收益保持觉察，这也是要求律师保持专业水准的通用型职业规范的一部分。

根据ABA对《职业行为示范规则》第1.6条（c）款的官方注释[18]和[19][4]，如果律师已采取合理措施预防，则对于客户代理相关信息未经授权的访问或无意、未经授权的披露，不构成对于（c）款的违反。在判断律师的措施是否为"合理措施"时，考虑因素包括但不限于以下五点：**信息的敏感性（sensitivity of the information）、未采取额外保障措施的情况下信息披露的可能性（likelihood of disclosure**

1.需说明的是，ABA的《职业行为示范规则》本身并无法律约束力，而是一套供各州采纳的示范规则。美国绝大部分州均在不同程度上以该示范规则为蓝本制定了本州的律师职业行为规则。因此，笔者理解，ABA的示范规则及其伦理意见（如文中引述的第477R号和第512号意见）在美国法律界具有广泛且权威的指导意义。
2.第 1.6 条（a）款的原文为：A lawyer shall not reveal information relating to the representation of a client unless the client gives informed consent, the disclosure is impliedly authorized in order to carry out the representation or the disclosure is permitted by paragraph (b).
3.第1.6条（c）款的原文为：A lawyer shall make reasonable efforts to prevent the inadvertent or unauthorized disclosure of, or unauthorized access to, information relating to the representation of a client.
4.https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_1_6_confidentiality_of_information/comment_on_rule_1_6/.

if additional safeguards are not employed）、采用额外保障措施的成本（cost of employing additional safeguards）、实施保障措施的困难程度（difficulty of implementing the safeguards）、保障措施对于律师代表客户能力的不利影响程度（extent to which the safeguards adversely affect the lawyer's ability to represent clients）。该官方注释也强调，如果信息传输方式本身已经提供了对于隐私保护的合理期待，本项责任并不要求律师采用特别的安全措施。

2017年，ABA伦理与职业责任常设委员会（Standing Committee on Ethics and Professional Responsibility）发布第477R号正式意见。该意见指出，《职业行为示范规则》并未因沟通方式的技术差异而设定更高或不同的保密义务，然而，在日新月异的技术环境中，律师应如何履行保密的核心职责，仍值得深思。

综上，ABA采取的"合理措施"标准，是基于具体事实的路径，要求建立一套流程，用于评估风险、识别并落实与之相适应的安全措施，验证其实施成效，并随新情况的发展持续优化。这种"过程论"的思路，为律师事务所根据具体情况灵活选择技术方案提供了空间。

3．针对生成式人工智能工具的进一步指引

2024年7月，ABA发布了专门针对生成式人工智能工具的第512号正式意见。就律师对于此类工具的使用，该意见重申了"合理措施"标准下的保密义务以及上述五项判断因素，并指出，由于现有各类生成式人工智能工具在"能否确保委托事项相关信息不被违规披露或访问"这一能力上存在差异，相关风险分析必须结合个案事实进行，具体取决于客户类型、案件性质、待完成工作以及所选用的生成式人工智能工具。

同时，意见指出，如使用自学习（self-learning）生成式人工智能工具，律师在输入客户代理信息前必须获得客户的知情同意[5]，如不涉及该等信息输入（如创意生成），则无需额外取得同意。

5.笔者理解，"自学习（self-learning）"在此并非笼统地指所有生成式人工智能工具，而是指某类工具会将律师输入的、与特定客户代理相关的信息，用于其自身的持续训练、优化或知识库/模型更新，从而使该等信息可能在后续其他请求的输出中，以直接或间接的形式暴露给本应无权获悉该信息的他人。正是这种潜在的"信息外溢"风险，构成了ABA第512号正式意见所强调的"在输入自学习人工智能工具前需事先取得客户知情同意"的逻辑基础。

总的来说，由于生成式人工智能技术发展迅速并存在不确定性，意见提出，所有律师都应当亲自阅读并理解其所用生成式人工智能工具的用户条款、隐私政策及相关合同文件，以弄清谁有权获取律师输入的信息；或者与已研读并分析过这些文件的同事或外部专家进行核实。为充分理解这些条款政策以及生成式人工智能工具处理信息的具体机制，律师还有必要咨询IT专业人员或网络安全专家。

4．国际律师事务所的人工智能部署实践

The American Lawyer于2024年初公布的调研结果显示，美国Am Law100律师事务所中，已有41家确认其已经开始使用生成式人工智能工具。

根据PwC公布的2024年英国律师事务所调研报告，英国前100名律师事务所中，接近90%已经实施或测试了生成式人工智能工具。

笔者根据公开信息整理了部分知名国际律师事务所采用生成式人工智能工具的情况，具体参见下表[6]：

律师事务所	生成式人工智能工具	类型
A&O Shearman（原Allen & Overy）[7]	Harvey	法律AI
Baker McKenzie[8]	Copilot for Microsoft 365	通用AI
Cleary Gottlieb[9]	Legora	法律AI
Clifford Chance[10]	Copilot for Microsoft 365	通用AI

6.此表为公开信息基础上的示意性梳理，不构成对各所全部技术采用情况的完整列举。
7.早在2023年，Allen & Overy已在官网宣布引入Harvey（https://www.aoshearman.com/en/news/ao-announces-exclusive-launch-partnership-with-harvey）；2025年，A&O Shearman在官网宣布与Harvey合作上线可执行复杂法律任务的系列智能体，这些智能体将在事务所内部实施，同时也向客户和其他律师事务所出售（https://www.aoshearman.com/en/news/ao-shearman-and-harvey-to-roll-out-agentic-ai-agents-targeting-complex-legal-workflows）。
8.Baker McKenzie在官网表示其进行了截至2024年法律行业最大的Microsoft Copilot部署。 https://www.bakermckenzie.com/en/newsroom/2025/06/most-innovative-global-law-firms-last-20-years-ft.
9.Cleary Gottlieb宣布与Legora达成战略合作，在事务所内上线Legora（https://legora.com/newsroom/cleary-gottlieb-announces-strategic-partnership-with-legora）。同时，Cleary Gottlieb收购了AI公司Springbok AI，并将其平台引入事务所（https://www.clearygottlieb.com/news-and-insights/news-listing/cleary-gottlieb-acquires-springbok-ai）。
10.Clifford Chance在官网宣布在该所全球范围内部署Copilot for Microsoft 365 and Viva Suite。https://www.cliffordchance.com/news/news/2024/02/clifford-chance-generative-ai-microsoft.html.

律师事务所	生成式人工智能工具	类型
DLA Piper[11]	CoCounsel	法律AI
Latham & Watkins[12]	Harvey	法律AI
Linklaters[13]	Legora	法律AI
Paul, Weiss[14]	Harvey（Workflow Builder）	法律AI（定制工作流工具）

值得注意的是，从公开信息来看，笔者未发现这些产品在常规商业模式下提供本地部署选项的说明。反而，Harvey、Copilot Enterprise等均基于企业级数据隔离架构。这意味着服务商虽然在云端处理数据，但在合同和技术上承诺不保留数据、不将数据用于模型训练[15]。这与大众使用的公有云上的AI产品存在本质区别。

5．一些启示

美国法律行业的实践，其价值不在于提供可以直接移植的技术方案，而在于向我们展示了一种解决保密义务与律师利用技术手段之间关系的方法。我们可以将其概括为以下三点：

首先，保密义务的本质是风险管理。关键不在于律师是否使用了第三方技术服务，而在于如何在法律框架通过合同、技术和管理手段

11.2023年，DLA Piper在官网宣布将使用Casetext（已被汤森路透收购）旗下CoCounsel。https://www.dlapiper.com/en/news/2023/03/dla-piper-to-utilize-cocounsel-the-groundbreaking-ai-legal-assistant-powered-by-openai-technology.
12.Latham & Watkins 在官网公布，事务所与Harvey签署企业级许可，将在全所范围内采用Harvey。https://www.lw.com/en/news/2025/08/latham-announces-firmwide-deployment-of-harvey.
13.Linklaters宣布在全所范围内落地Legora。https://legora.com/newsroom/linklaters-announces-firmwide-rollout-of-legora.
14.Paul, Weiss在官网公布，其成为首家上线Harvey工作流定制工具Workflow Builder的律师事务所，并将作为Harvey在该工具设计方面的核心合作伙伴。https://www.paulweiss.com/insights/firm-news/paul-weiss-partners-with-harvey-ai-on-new-ai-workflows-innovation.
15.Harvey：“Neither Harvey nor its AI subprocessors (such as Microsoft and OpenAI) use your data to train AI models. Harvey's AI subprocessors do not retain your data and only process your data ephemerally...”，参见：trust.harvey.ai.
Microsoft 365 Copilot：“Your data isn't used to train foundation models”，参见：https://learn.microsoft-.com/en-us/copilot/microsoft-365/enterprise-data-protection.
Thomson Reuters CoCounsel：“Thomson Reuters AI third-party partners, such as OpenAI and Google, are contractually prohibited from using any customer data to train their models.”参见：https://www.thomsonreuters.com/en/cocounsel.
Legora：“Your confidential data remains secure and private to you. Legora will not use your data to train or fine tune any AI models.”，参见https://legora.com/security.

将风险控制在可接受的范围内。

其次，合理措施是一个动态的、具体情况具体分析的标准。不同敏感度的信息、不同的业务场景、不同的技术能力，需要匹配不同程度的保护措施。

第三，对技术服务商的尽职调查和合同约束是风险管理的关键环节。

我们也需要认识到，中美两国在法律体系、监管环境、技术基础设施等方面存在差异，美国的做法不可能被简单复制。但美国律师行业发展出的如何在严格的职业义务约束下，通过系统性的风险评估和分级管理，实现对新技术的合理利用，或许值得我们思考和借鉴。

当然，对于中国律师事务所来说，所有的技术选型都必须在《保守国家秘密法》《律师法》《数据安全法》《个人信息保护法》《数据出境安全评估办法》等法律法规的框架内进行。

003 > 技术路径的现实考量与可行选择探讨

随着DeepSeek、千问等优秀国产大语言模型的开源，与本地部署闭源模型相比，组织以更低成本进行本地化部署模型成为可能。选择本地化部署具有充分的合理性，即将数据和计算都保留在组织内部，从物理上隔离外部访问。然而，笔者认为，生成式人工智能的技术特性，也为本地化部署及后续能否有效使用带来实际的挑战。

需要认识到的是，当前生成式人工智能技术的发展和应用生态主要构建在云端。这不仅是因为云服务提供了必要的技术基础设施，更因为大参数模型高并发、不延迟运行所需的算力规模巨大。具体而言，如果仅在本地部署大语言模型（如DeepSeek），组织会面临诸多限制。例如，模型无法进行联网搜索，只能输出截至模型训练日的信息；如需实现联网搜索、文档处理、多模态处理等功能，需要额外开发大量配套系统，而云端通常提供较成熟的可用组件。

同时，仅部署大语言模型本身并不能直接产生业务价值。要让AI真正服务于法律工作，需要开发相应的应用系统、集成现有业务流

程、建立数据管理机制等。这些配套工作的复杂度和成本，往往超过模型部署本身。

在成本方面，本地化部署需要硬件采购、技术栈搭建和专业运维团队，初始投入和持续成本较高。更重要的是，人工智能技术快速演进，本地部署的升级成本和时间成本都相对较高。

笔者认为，如果仅仅本地部署模型而缺乏成熟的配套应用，可能会导致投入产出比不佳，且难以满足律师日益增长的智能化作业需求。在满足监管与客户要求的前提下，基于数据的不同类型，采用物理隔离或是位于境内的合规且受控的企业级云服务，或许是中国律师事务所平衡效率与合规的务实路径。在采用云服务的情况下，应通过合同与技术隔离明确禁止将客户代理信息用于训练第三方模型，并落实访问控制与日志审计。

但是，我们也必须强调，本地部署的挑战主要是技术路径带来的客观局限。在律师行业缺乏明确监管指引、对云服务信任机制尚未完全建立的情况下，如律师事务所选择看似最安全的本地化部署方案，是可以充分理解的谨慎负责态度。面对这些挑战，技术手段只是风险管理的一个维度。中国律师行业的人工智能应用最终会走出什么样的道路，还有待时间给出答案，但可以确定的是，这需要技术理解、风险评估、制度建设、行业协同的共同推进。

004 > 结语

保密义务是律师职业的基石，这一点在人工智能时代不会也不应改变。保密义务的本质是风险管理，且应随着技术的发展赋予其相匹配的边界。本地化部署、受控企业级云环境、公有云，各有其适用场景，关键在于建立与风险相匹配的管理机制。

律师作为承担严格职业责任的群体，在人工智能应用探索中的审慎态度是必要的，也是对客户负责的体现。但审慎不等于停滞。笔者认为，对于律所而言，人工智能部署应重点考量的是如何在法律法规和监管要求的框架下谨慎推进实施的问题。

这条路径的探索需要监管指引、行业共识和技术创新的协同推进。相信中国律师行业终将在严守保密义务与合理利用技术之间，找到属于自己的平衡之道。

（本文系笔者作为一名律师在人工智能浪潮中的个体观察与思考，不代表本人所在律师事务所的观点，亦不构成正式法律意见。）

韩梅
部门负责人
知识管理部
北京办公室
+86 10 5957 2294
hanmei@zhonglun.com

附录

生成式人工智能产品合规义务清单V3.0

generative ai product compliance obligation list v3.0

参考法律法规、标准、指南[1]：

- 《网络安全法》
- 《数据安全法》
- 《个人信息保护法》
- 《网络数据安全管理条例》（《网数条例》）
- 《互联网信息服务算法推荐管理规定》（《算法推荐管理规定》）
- 《互联网信息服务深度合成管理规定》（《深度合成管理规定》）
- 《生成式人工智能服务管理暂行办法》（《AIGC管理办法》）
- 《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》（《安全评估规定》）
- 《互联网用户账号信息管理规定（2022）》（《账号管理规定》）
- 《网络信息内容生态治理规定》（《内容生态治理规定》）
- 《科技伦理审查办法（试行）》
- 《人工智能生成合成内容标识办法》（《标识办法》）
- 《 网络安全技术 生成式人工智能服务安全基本要求》（GB/T 45654-2025）
- 《网络安全技术 生成式人工智能预训练和优化训练数据安全规范》（GB/T 45652-2025）
- 《网络安全技术 生成式人工智能数据标注安全规范》（GB/T 45674-2025）
- 《网络安全技术 人工智能生成合成内容标识方法》（GB 45438-2025）

适用范围：

1)生成式人工智能技术是指具有文本、图片、音频、视频等内容生成能力的模型及相关技术。

2)利用生成式人工智能技术向我国境内公众提供生成文本、图片、音频、视频等内容的服务。

1.截止至2025年10月发布的法律法规以及正式发布的相关标准等。

3)行业组织、企业、教育和科研机构、公共文化机构、有关专业机构等研发、应用生成式人工智能技术，未向境内公众提供生成式人工智能服务的，不适用《办法》中的合规要求。

编者注：

1)本清单发布于2025年10月更新，仅用于相关企业的合规参考，不作为正式的法律意见。

001> 开发&训练阶段

合规要点		具体合规义务	法律依据
训练数据管理	数据来源安全	➢ 制定人工智能预训练和优化训练数据的安全管理策略 ➢ 训练数据应具有合法来源 ➢ 从外部数据源收集的预训练数据，应记录数据收集所涉及的数据来源 ➢ 对采集数据来源进行随机抽样安全评估，包含违法不良信息超过5%的不应采集 ➢ 对每个来源已采集数据进行随机抽样行安全核验，包含违法不良信息超过5%的不应使用 ➢ 使用开源语料，应遵循该数据来源的开源许可协议或取得相关授权文件 ➢ 使用自采训练，应具有采集记录，不应采集他人已明确不可采集的数据 ➢ 使用商业训练数据时，应具备法律效力的交易合同、合作协议等，交易方或合作方应提供数据承诺以及相关证明材料 ➢ 将使用者输入信息用作训练数据时，应具有使用者授权记录 ➢ 收集生成式人工智能服务生成的数据时，应记录所使用生成式人工智能服务的提供商、版本、获取时间、数据标识等信息 ➢ 使用生成式人工智能生成内容作为训练数据时，应建立幻觉风险评估机制，识别并处置误导模型的错误知识	《AIGC管理办法》第七条 《GB/T 45654-2025》第4.1条 《GB/T 45652-2025》第4a）条 《GB/T 45652-2025》第5.1、6.1、6.3条

合规要点	具体合规义务	法律依据
数据内容安全	➢ 定期对预训练和优化训练数据开展安全评估 ➢ 预训练和优化训练数据进行安全检测，修复或过滤被投毒数据，进行真实性评估 ➢ 采取措施降低生成式人工智能被诱导生成安全风险内容的可能性，包括但不限于充分过滤已识别含有安全风险内容的数据样本等。 ➢ 采取关键词、分类模型、人工抽检等方式充分过滤语料中违法不良信息 ➢ 涉及个人信息（敏感个人信息）的，应当取得个人同意（单独同意）或者符合法律、行政法规规定的其他情形 ➢ 训练数据涉及知识产权的，具备训练数据知识产权管理策略和规则，并明确负责人，不得侵害他人依法享有的知识产权，建立知识产权投诉举报渠道，重点识别著作权侵权风险，在用户服务协议中向使用者告知使用生成内容的知识产权相关风险，并约定相关责任与义务	《个人信息保护法》第十三条 《AIGC管理办法》第七条 《GB/T 45654-2025》第4.2条 《GB/T 45652-2025》第4l）、4n）、4o）、5.3条
数据质量把控	➢ 增强训练数据的真实性、准确性（例如，运用数据清洗和预处理技术，排除训练数据中的噪声和偏差） ➢ 增强训练数据的客观性、多样性（例如，通过搜集来自不同来源、不同领域和不同背景的数据，确保训练数据集具有广泛的覆盖范围，减少因为偏向性而导致的结果失衡问题） ➢ 算法设计、训练数据选择、模型生成和优化过程中采取有效措施防止产生民族、信仰、国别、地域、性别、年龄、职业、健康等歧视	《AIGC管理办法》第七条

合规要点	具体合规义务	法律依据
数据标注规范	➢ 对标注人员进行培训和考核，给予合格者标注上岗资格，并有定期重新培训考核以及必要时暂停或取消标注上岗资格的机制 ➢ 同一标注任务下，标注执行人员和标注审核人员不应由同一人员担任 ➢ 应对各数据标注人员任务分配过程进行记录，并留存记录文档 ➢ 制定具体、可操作的安全性及功能性标注规则，至少包括标注目标、标注方法和质量指标等内容 ➢ 对功能性数据标注以及安全性数据标注分别制定标注规则 ➢ 标注内容需经过审核，对于功能性标注每一批语料应进行人工抽检，对安全性标注每一条标注语料至少一名人员审核 ➢ 对安全性标注数据进行隔离存储	《AIGC管理办法》第八条 《GB/T 45654-2025》第4.3条 《GB/T 45674-2025》第6、7.2条
数据安全措施	➢ 采取必要措施保障训练数据安全（例如访问控制、加密等技术措施） ➢ 加强数据训练活动和训练数据处理活动的安全管理	《网数条例》第十九条 《深度合成管理规定》第十四条

合规要点	具体合规义务	法律依据
模型安全要求	➢ 训练过程中，将生成内容安全性作为评价主要考虑指标 ➢ 定期对所使用的开发框架、代码等进行安全审计，关注开源框架安全以及漏洞相关问题，定期进行后门存在性检测 ➢ 模型应定期进行监测和优化，确保和提高生成内容的安全性、准确性和可靠性 ➢ 生成内容安全性方面，应保证模型生成内容合格率不低于90% ➢ 生成内容价值正确：生成内容不得违背社会主义核心价值观，不得生成法律、行政法规禁止的内容 ➢ 不得生成歧视内容：应采取有效措施防止产生民族、信仰、国别、地域、性别、年龄、职业、健康等歧视内容 ➢ 建立常态化监测测评手段以及模型应急管理措施，对监测测评发现的提供服务过程中的安全问题，及时处置并通过针对性的指令微调、强化学习等方式优化模型	《AIGC管理办法》第四条 《算法推荐管理规定》第九条 《深度合成管理规定》第十条 《GB/T 45654-2025》第5条

002 > 对外提供技术阶段

合规要点	具体合规义务	法律依据
安全评估	➢ 深度合成技术支持者提供生成或者编辑人脸、人声等生物识别信息的以及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模型的，应当开展安全评估	《深度合成管理规定》第十五条
算法备案	➢ 深度合成技术支持者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履行备案手续 ➢ 深度合成技术支持者应敦促、协助调用其技术的深度合成服务提供者履行备案登记义务	《深度合成管理规定》第十九条 《算法推荐管理规定》第二十四条
备案编号公示	➢ 深度合成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	《深度合成管理规定》第十九条
信息安全	➢ 对外提供技术服务应根据网络安全法、数据安全法等履行基本的信息安全义务	《深度合成管理规定》第七条
科技伦理审查	➢ 设立科技伦理（审查）委员会 ➢ 开展人工智能科技伦理风险评估 ➢ 制定科技伦理风险评估办法 ➢ 建立需要开展专家复核的科技活动清单制度 ➢ 制定科技伦理应急审查制度	《科技伦理审查办法（试行）》第五、九条

003 > 对外提供服务阶段

合规要点	具体合规义务	法律依据
算法备案	➢ 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履行备案手续	《深度合成管理规定》第十九条 《算法推荐管理规定》第二十四条
大模型备案	➢ 提供具有舆论属性或者社会动员能力的生成式人工智能服务的，应在对外提供服务前完成大模型备案 ➢ 直接调用已备案模型的API并面向境内公众提供具有舆论属性或者社会动员能力的生成式人工智能服务的，应进行大模型调用登记	《AIGC管理办法》第十七条
内部控制制度	➢ 算法机制机理审核机制 ➢ 科技伦理审查机制 ➢ 用户账号管理相关制度、用户账号信用管理体系 ➢ 信息发布（及跟帖评论）审核机制 ➢ 数据安全和个人信息保护制度 ➢ 投诉举报处理机制 ➢ 反电信网络诈骗制度 ➢ 安全评估监测机制 ➢ 安全事件应急处置等管理制度和技术措施	《网数条例》第二十条 《深度合成管理规定》第七、十五条 《算法推荐管理规定》第八条 《AIGC管理办法》第十五条
内部组织人员	➢ 科技伦理（审查）委员会 ➢ 网络信息内容生态治理负责人 ➢ 算法安全责任人 ➢ 知识产权负责人 ➢ 数据安全和个人信息保护机构	《科技伦理审查办法（试行）》第五条 《内容生态治理规定》第九条
服务透明度	➢ 在网站首页等显著位置、API说明文档明确并公开服务的适用人群、场合、用途，同时公开基础模型使用情况 ➢ 在网站首页、服务协议、API说明文档等便于查看的位置向使用者公开服务的局限性、所使用模型的概要信息以及个人信息及其在服务中的用途 ➢ 算法信息公示，包括基本原理、目的意图和主要运行机制 ➢ 算法备案编号及算法备案公示信息链接（网站、应用程序等的显著位置）	《算法推荐管理规定》第十六条 《AIGC管理办法》第十条 《GB/T 45654-2025》第6.2条

合规要点	具体合规义务	法律依据
个人信息处理合规	➢ 遵循《个人信息保护法》等相关法律法规要求 ➢ 生物识别信息单独同意：提供人脸、人声等生物识别信息编辑功能的，应提示服务使用者告知被编辑的个人并取得其单独同意 ➢ 人工干预机制：采用个性化算法推荐技术推送信息的，应当建立健全人工干预机制 ➢ 算法推荐退出选项：提供不针对其个人特征的选项，或者关闭算法推荐服务的选项 ➢ 用户标签编辑功能：提供选择或者删除用于算法推荐服务的针对其个人特征的用户标签的功能 ➢ 禁止差别待遇：不得利用算法在交易价格等交易条件上实施不合理的差别待遇	《AIGC管理办法》第九、十一条 《深度合成管理规定》第十四条 《账号管理规定》第十六条 《算法推荐管理规定》第十七、二十一条
服务稳定、持续	➢ 训练环境和推理环境隔离，避免数据泄露和不当访问 ➢ 持续监测模型输入内容，防范恶意输入攻击，如注入攻击、数据窃取、对抗攻击等 ➢ 制定在模型更新、升级时的安全管理策略，在模型重要更新、升级后，组织安全评估 ➢ 建立数据、模型、框架、工具等的备份机制以及恢复策略，重点确保业务连续性	《AIGC管理办法》第十三条 《GB/T 45654-2025》第5.3、5.4、5.5、6.6条
投诉、举报机制	➢ 设置接受公众或使用者便捷的投诉、举报入口（例如在线平台、电子邮件、交互窗口、热线电话、短信等多种形式） ➢ 通过平台规则、服务协议或者产品页面等方式公布处理流程和反馈时限 ➢ 及时受理、处理公众投诉举报并反馈处理结果	《AIGC管理办法》第十五条 《深度合成管理规定》第十二条 《内容生态治理规定》第十六条 《安全评估规定》第五条 《GB/T 45654-2025》第6.4条

合规要点		具体合规义务	法律依据
生成结果标识	隐式标识	➢ 企业应对生成的内容采取技术措施添加不影响用户使用的标识，并依法保存相关日志信息 ➢ 企业应在生成合成内容的文件元数据中添加隐式标识，包含生成合成内容属性信息、服务提供者名称或编码、内容编号等制作要素信息，具体包括： 1）生成合成标签要素：内容的人工智能生成合成属性信息 2）生成合成服务提供者要素：生成合成服务提供者的名称或编码 3）内容制作编号要素：生成合成服务提供者对该内容的唯一编号 4）内容传播服务提供者要素：内容传播服务提供者的名称或编码 5）内容传播编号要素：内容传播服务提供者对该内容的唯一编号 ➢ 人工智能生成合成的内容文件中，应仅保留一份文件元数据隐式标识	《AIGC管理办法》第十二条 《深度合成管理规定》第十六、十七条 《标识办法》第三、五条 《GB 45438-2025》第6条
	显式标识	**以下场景除添加隐式标识外，还需添加显式标识：** ➢ 智能对话、智能写作等模拟自然人进行文本的生成或者编辑服务 ➢ 合成人声、仿声等语音生成或者显著改变个人身份特征的编辑服务 ➢ 人脸生成、人脸替换、人脸操控、姿态操控等人物图像、视频生成或者显著改变个人身份特征的编辑服务 ➢ 沉浸式拟真场景等生成或者编辑服务 ➢ 文生图片等图片内容生成服务 ➢ 音乐创作等音频内容生成服务 ➢ 文生视频、图生视频等视频内容生成服务	《AIGC管理办法》第十二条 《深度合成管理规定》第十六、十七条 《标识办法》第三、四条 《GB 45438-2025》第5条、附录B

合规要点	具体合规义务	法律依据
	➢ 其他具有生成或者显著改变信息内容功能的服务 **显式标识必须同时包含人工智能要素和生成合成要素：** ➢ 人工智能要素：包含“人工智能”或“AI”，表明使用人工智能技术 ➢生成合成要素：包含“生成”和/或“合成”，表明内容制作方式为生成和/或合成	
	文本内容标识要求，在文本的起始、末尾或者中间适当位置添加文字提示或者通用符号提示等标识，或者在交互场景界面、文字周边添加显著的提示标识： ➢ 文本内容显式标识应采用文字或角标形式 ➢ 在文本的起始、末尾、中间适当的一个或多个位置 ➢ 文本内容显式标识使用的字型和颜色应清晰可辨	
	图片内容标识要求，在图片的适当位置添加显著的提示标识： ➢ 图片内容显式标识应采用文字提示 ➢ 显式标识应位于图片的边或角，文字高度应不低于画面最短边长度的5% ➢ 图片内容显式标识使用的字型应清晰可辨	
	音频内容标识要求，在音频的起始、末尾或者中间适当位置添加语音提示或者音频节奏提示等标识，或者在交互场景界面中添加显著的提示标识： ➢ 音频内容显式标识应采用语音标识或音频节奏标识 ➢ 音频节奏标识应为“短长 短短”的节奏，语音标识应使用正常语速（约120~160字/分钟） ➢ 在音频的起始、末尾、中间适当的一个或多个位置 ➢ 节奏标识应清晰可辨	

合规要点	具体合规义务	法律依据
	视频内容标识要求： ➢ 显式标识应采用文字提示 ➢ 显式标识应位于视频起始画面，可位于视频末尾和中间适当位置 ➢ 显式标识应位于视频画面的边或角，显式标识的文字高度应不低于画面最短边长度的5%	
	虚拟场景标识要求，在起始画面的适当位置添加显著的提示标识，可以在虚拟场景持续服务过程中的适当位置添加显著的提示标识： ➢ 显式标识应采用文字提示 ➢ 应位于虚拟场景起始画面，可位于虚拟场景持续服务过程中的适当位置 ➢ 位于虚拟场景起始画面的虚拟场景显式标识应位于画面的边或角 ➢ 虚拟场景显式标识使用的字型应清晰可辨 ➢ 文字高度不应低于画面最短边长度的5%	
	交互界面标识要求： ➢ 显式标识应采用文字提示 ➢ 显式标识应在内容展示区域附近持续显示、在内容展示区域的背景均匀展示，或在音频、视频交互区域附近或界面顶部、底部持续显示的一种或多种方式 ➢ 交互场景界面显式标识使用的字型和颜色应清晰可辨	
配套义务	➢ 在用户服务协议中明确说明生成合成内容标识的方法、样式等规范内容，并提示用户仔细阅读并理解相关的标识管理要求 ➢ 用户需要服务提供者提供没有添加显式标识的生成合成内容的，可通过用户协议明确用户的标识义务和使用责任后提供，相关日志留存不少于六个月 ➢ 加强标识信息共享，为防范打击相关违法犯罪活动提供支持和帮助	《标识办法》第八、九、十二条

合规要点	具体合规义务	法律依据
违法和不良信息特征库	➢ 建立健全用于识别违法和不良信息的特征库； ➢ 记录并留存相关网络日志	《算法推荐管理规定》第九条 《深度合成管理规定》第十条
违法违规处置	➢ 采取关键词、分类模型等方式对使用者输入信息进行检测 ➢ 设置并向使用者公示以下规则：在使用者连续多次输入违法不良信息或一天内累计输入违法不良信息达到一定次数时，采取暂停提供服务等处置措施 ➢ 依法依约对使用生成式人工智能服务从事违法活动的用户采取警示、限制功能、暂停或者终止向其提供服务等处置措施 ➢ 建立健全辟谣机制，监控虚假信息的制作、复制、发布和传播 ➢ 设置拒答机制，拒绝回答明显偏激以及明显诱导生成违法不良信息的问题 ➢ 设置监看人员，并及时根据监看情况提高生成内容质量及安全 ➢ 优化整改：采取模型优化训练等措施进行整改 ➢ 报告义务：向网信部门和有关主管部门报告 ➢ 处置记录保存：企业应对违法信息、违法活动内容以及处置措施进行记录 注：违法违规内容的范围参见《内容生态治理规定》第六条	《AIGC管理办法》第十四条 《算法推荐管理规定》第九条 《深度合成管理规定》第十条 《内容生态治理规定》第十条 《安全评估规定》第五条 《用户账号管理》第十七条 《GB/T 45654-2025》第6.5条
信息内容治理	➢ 建立完善人工干预和用户自主选择机制，在首页首屏、热搜、精选、榜单类、弹窗、联想词、皮肤、推荐区等重点环节积极呈现符合主流价值导向的信息 ➢ 企业应当编制网络信息内容生态治理工作年度报告，年度报告应当包括：网络信息内容生态治理工作情况、网络信息内容生态治理负责人履职情况、社会评价情况等内容	《算法推荐管理规定》第十一条 《内容生态治理规定》第十一、十七条
端侧模型服务	➢ 在使用者首次使用服务时通过官方途径进行激活，并在设备联网时推送安全策略更新 ➢ 具备端侧安全模块：应利用关键词库等技术对生成内容进行安全审核，收集并留存安全日志，并支持设备联网时上传日志或支持端侧本地导出日志；应在设备联网时定期更新关键词库以及相关安全配置 ➢ 具备模型更新机制：发现模型安全漏洞时，应及时对安全漏洞进行修复，例如推送安全补丁到端侧等；当模型有重大更新时，应针对长时间未更新的端侧使用者，提供多次提醒和预警	《GB/T 45654-2025》第6.7条

合规要点	具体合规义务	法律依据
外部监管证照资质	➢ ICP备案/许可 ➢ 公安联网备案 ➢ App、小程序备案 ➢ 互联网新闻信息服务许可（如需） ➢ 互联网宗教信息服务学（如需） ➢ 网络文化经营许可（如需） ➢ 网络出版服务许可（如需） ➢ 信息网络传播视听节目许可或备案（如需）	《AIGC管理办法》第十七条 《算法推荐管理规定》第二十四条等
实名认证	企业应当基于移动电话号码、身份证件号码、统一社会信用代码或者国家网络身份认证公共服务等方式对使用者进行真实身份信息认证	《深度合成管理规定》第九条 《账号管理规定》第九条

合规要点		具体合规义务	法律依据
账号信息管理	不予注册	➢ 账号含有法律、行政法规和国家有关规定禁止情形（《内容生态治理规定》第六条、第七条、《账号管理规定》第八条） ➢ 被依法依约关闭的账号重新注册	《AIGC管理办法》第十四条 《算法推荐管理规定》第九条 《账号管理规定》第六、七、八、十、十五条 《内容生态治理规定》第六、七条
	从严审核	➢ 对账号信息中含有“中国”“中华”等内容，或者含有党旗、党徽、国旗等党和国家象征和标志的 ➢ 与被依法关闭账号关联度高的新账号	
	必须核验	➢ 个人用户账号的职业信息（如需填写）是否与个人真实职业信息相一致 ➢ 机构用户账号是否与其实际机构名称、标识等相一致，与机构性质、经营范围和所属行业类型等相符合 ➢ 申请注册从事经济、教育、医疗卫生、司法等领域信息内容生产的账号，企业还需核验其提供服务资质、职业资格、专业背景等相关材料，并在账号信息中加注专门标识	
	动态核验	企业应当适时核验存量账号信息，发现不符合《账号管理规定》要求的，应当暂停提供服务并通知用户限期改正；拒不改正的，应当终止提供服务	
账号信息展示	个人账号	在用户账号信息页面展示合理范围内的IP地址归属地信息	《账号管理规定》第十二条
	公众账号	在公众账号信息页面展示公众账号的运营主体、注册运营地址、内容生产类别、统一社会信用代码、有效联系方式、IP地址归属地等信息	《账号管理规定》第十三条

合规要点	具体合规义务	法律依据
信息保存	➢ 注册信息保存：企业应对用户的注册信息采取留存措施 ➢ 使用信息保存：对用户的账号、操作时间、操作类型、网络源地址和目标地址、网络源端口、客户端硬件特征等日志信息，以及用户发布信息记录的留存措施	《安全评估规定》第五条
协议及平台规则	➢ 制定服务协议，明确与用户之间的权利义务，提示用户履行信息安全义务 ➢ 遵循《个人信息保护法》等相关法律法规要求，制定隐私政策等个人信息保护规则 ➢ 为使用者提供关闭其输入信息用于训练的方式，采用选项方式时使用者从服务主界面开始到达该选项所需操作不超过4次点击 ➢ 公布投诉、举报规则，包括入口、处理流程、反馈时限，及时受理、处理和反馈处理结果 ➢ 制定信息内容使用规则、知识产权保护规则等平台管理规则	《AIGC管理办法》第九、十一条 《深度合成管理规定》第八、十四条 《账号管理规定》第十六条 《生态治理规定》第十五条 《算法推荐管理规定》第七条 《GB/T 45654-2025》第6.3、6.4条

合规要点		具体合规义务	法律依据
特殊主体保护	未成年人	➢ 遵循《未成年人保护法》《未成年人网络保护条例》《儿童个人信息网络保护规定》等相关法律法规要求 ➢ 开发未成年人模式应用、提供适合未成年人特点的服务 ➢ 不得推送可能引发未成年人模仿不安全行为和违反社会公德行为、诱导未成年人不良嗜好等可能影响未成年人身心健康的信息 ➢ 采取有效措施防范未成年人用户过度依赖或者沉迷网络 ➢ 针对未成年人使用的生成式人工智能服务，应允许监护人设置防沉迷措施，且不向未成年人提供超出其行为能力的付费服务，积极展示有益未成年人身心健康的内容 ➢ 服务不适用未成年人的，应采取技术或管理措施防止未成年人使用	《AIGC管理办法》第十条 《算法推荐管理规定》第十八条 《内容生态治理规定》第十三条 《GB/T 45654-2025》第6.1条
	老年人	➢ 充分考虑老年人出行、就医、消费、办事等特殊需求提供智能化适老服务 ➢ 开展涉电信网络诈骗信息的监测、识别和处置	《算法推荐管理规定》第十九条
	劳动者	➢ 企业基于算法向劳动者提供工作调度服务的，应当保护劳动者取得劳动报酬、休息休假等合法权益，建立完善平台订单分配、报酬构成及支付、工作时间、奖惩等相关算法	《算法推荐管理规定》第二十条

合规要点	具体合规义务	法律依据
广告审查	➢ 遵循《互联网广告管理办法》等相关法律法规要求 ➢ 对平台设置的广告位和在平台展示的广告内容进行审核巡查 ➢ 利用算法推荐等方式发布互联网广告的，应当将其算法推荐服务相关规则、广告投放记录等记入广告档案	《内容生态治理规定》第十四条
APP（小程序）用户权益保障	➢ 产品开屏和弹窗信息窗口提供清晰有效的关闭按钮，保证用户可以便捷关闭；不得频繁弹窗干扰用户正常使用，或利用"全屏热力图"、高灵敏度"摇一摇"等易造成误触发的方式诱导用户操作 ➢ 清晰明示产品功能权益及资费等内容，显著提示开通会员、收费等附加条件 ➢ 在非服务所必需或无合理场景下，不得自启动和关联启动其他APP，或进行唤醒、调用、更新等行为 ➢ 遵守《个人信息保护法》等个人信息保护规定，坚持合法正当必要原则、明示个人信息处理规则、合理申请使用权限等 ➢ 妥善处理用户投诉 ➢ 建立SDK目录、建立个人信息保护"双清单"	《个人信息保护法》 《AIGC管理办法》第四条 《深度合成管理规定》第六条 《算法推荐管理规定》第六条
配合监管	➢ 配合网信部门和电信、公安、市场监管等有关部门开展安全评估和监督检查工作，并提供必要的技术、数据等支持和协助 ➢ 配合主管部门监督检查工作，按要求对训练数据来源、规模、类型、标注规则、算法机制机理等予以说明 ➢ 为公安机关、国家安全机关依法维护国家安全和查处违法犯罪提供技术、数据支持和协助	《算法推荐管理规定》第二十八条 《AIGC管理办法》第十九条 《安全评估规定》第五条

蔡鹏
合伙人
知识产权部
北京办公室
+86 10 5087 2786
caipeng@zhonglun.com

中伦研究院出品

www.zhonglun.com