中倫
ZHONG LUN

[人工智能2.0]

全景透视AIGC的法律挑战与合规路径

ARTIFICIAL INTELLIGENCE 2.0:

LEGAL & ETHICAL CHALLENGES OF AIGC

COMPREHENSIVE ANALYSIS & KEY STRATEGIES

法律总览丨AIGC的六大法律挑战

知识产权丨合理使用、可版权性与侵权责任

域外观察丨欧美人工智能知产合规

数据合规丨网络安全、数据合规与跨境合规

AIGC监管丨算法备案·安全评估·资质证照

侵权责任丨反不正当竞争与侵权责任

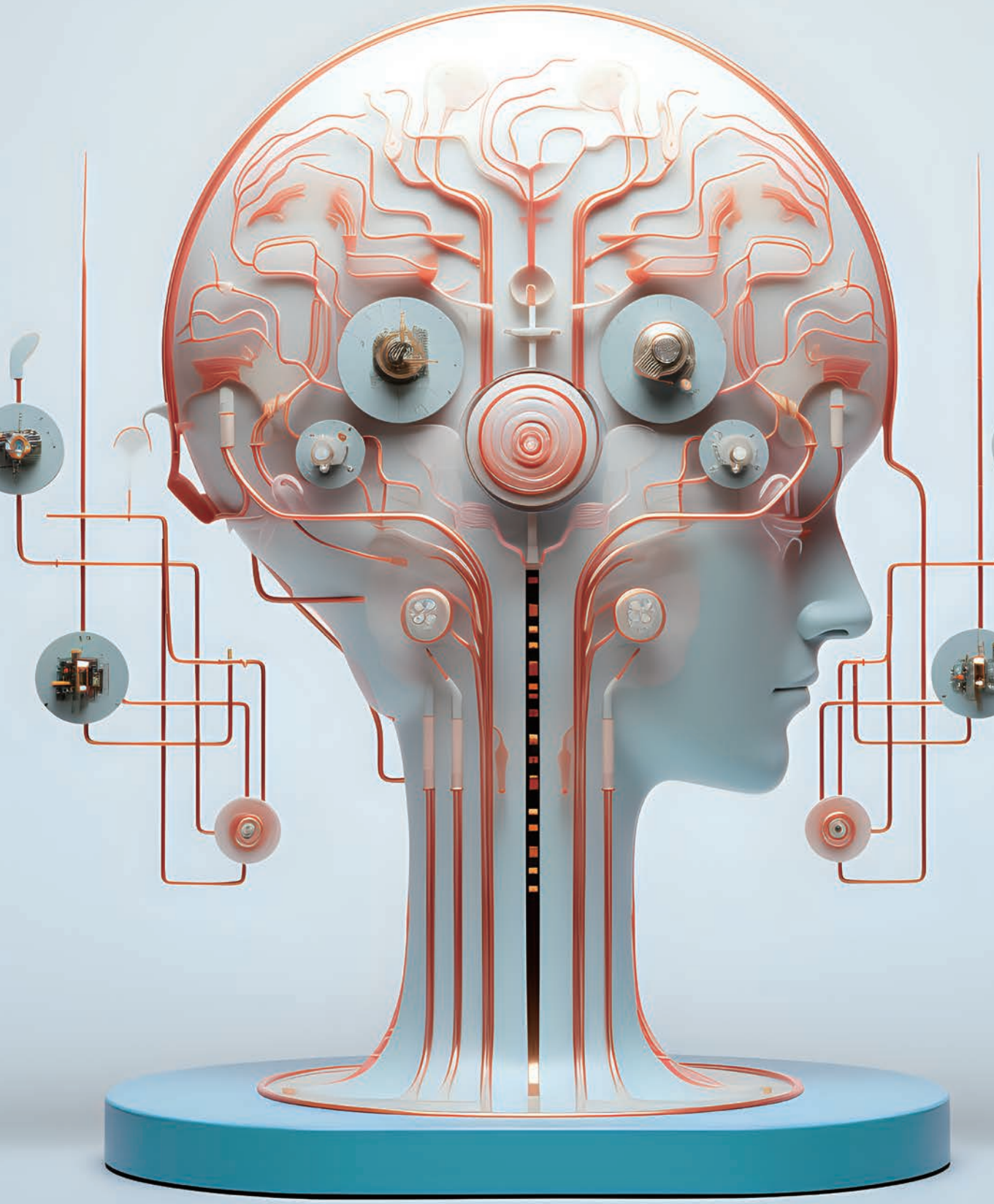
AI伦理丨伦理治理

ChatGPT
科技与法律
最新报告

中伦研究院出品

CONTENTS 目录

CONTENTS 目录



[第肆篇章]

[第伍篇章]

[第陆篇章]

PREFACE 前言

在这个日新月异、充满挑战与机遇的时代，人工智能已经深入到我们生活的各个角落。从聊天机器人的智慧对话，到艺术作品的创作灵感，再到企业决策的精准预测，AIGC正在以前所未有的方式改变着我们的世界。然而，随着AIGC的飞速发展，一系列知识产权、数据合规、政府监管、侵权责任和伦理等问题也随之涌现，给合规能力带来了严峻挑战。

例如，当"ChatGPT"这样的AI技术遇上法律的缰绳，会激发出怎样的火花？当AIGC的生成内容涉及到知识产权侵权，企业又该如何应对？当AIGC涉及数据跨境，企业又该如何保证其合规性？

在这本文集中，我们将从AIGC的法律问题概览出发，探讨知识产权的挑战与合规，透视数据合规的问题与合规路径，研究AIGC的监管合规挑战与应对，理解侵权责任的新形态及法律规制，思考AI伦理的问题与治理，并在最后奉上一份详尽的AIGC应用的合规指引清单。

在知识产权部分，我们将全景透视生成式人工智能在知识产权方面的挑战与合规；讨论AIGC生成内容的可版权性、权利归属与使用限制，以及侵权责任的承担；探讨涉美欧人工智能业务的知识产权合规要求、变化趋势及应对建议。

在数据合规部分，我们将以独特的方法论，分析人工智能生命周期的网

络安全与数据合规风险，包括在模型开发训练、应用运行、模型优化等阶段，回应数据合规挑战，并探讨AIGC数据跨境的法律监管和合规路径。

在监管方面，我们在《生成式人工智能服务管理暂行办法》的基础上；对境内提供AIGC服务的主要监管框架进行梳理，包括算法监管、安全评估、内容监管、增值电信监管和科技伦理监管，并对焦点问题进行探析。此外，我们还将给出AIGC产品的合规上市之路，包括算法备案与安全评估、资质证照等方面的内容。

在法律责任方面，我们将讨论机器学习作品的类型化及其著作权侵权责任，以及涉数据、算法等技术的新型不正当竞争行为及法律规制。同时，我们还将探讨人工智能服务提供者的归责原则。

生成式AI可能会创造出令人不安的内容，或者侵犯个人隐私。我们应该如何处理这些问题？因此最后，在AI伦理方面，我们将针对AI伦理所涉及的法律进行梳理，讨论生成式AI伦理治理的原则和立场。

我们希望，这份文集能够帮助企业在AIGC的世界里游刃有余，避免"假作真时真亦假"的困境，让企业在创新的道路上更加从容自信。

在这个充满变革与机遇的时代，让我们一起探索AIGC的法律与伦理挑战，共同迈向更加美好的未来！

A I

CHAPTER

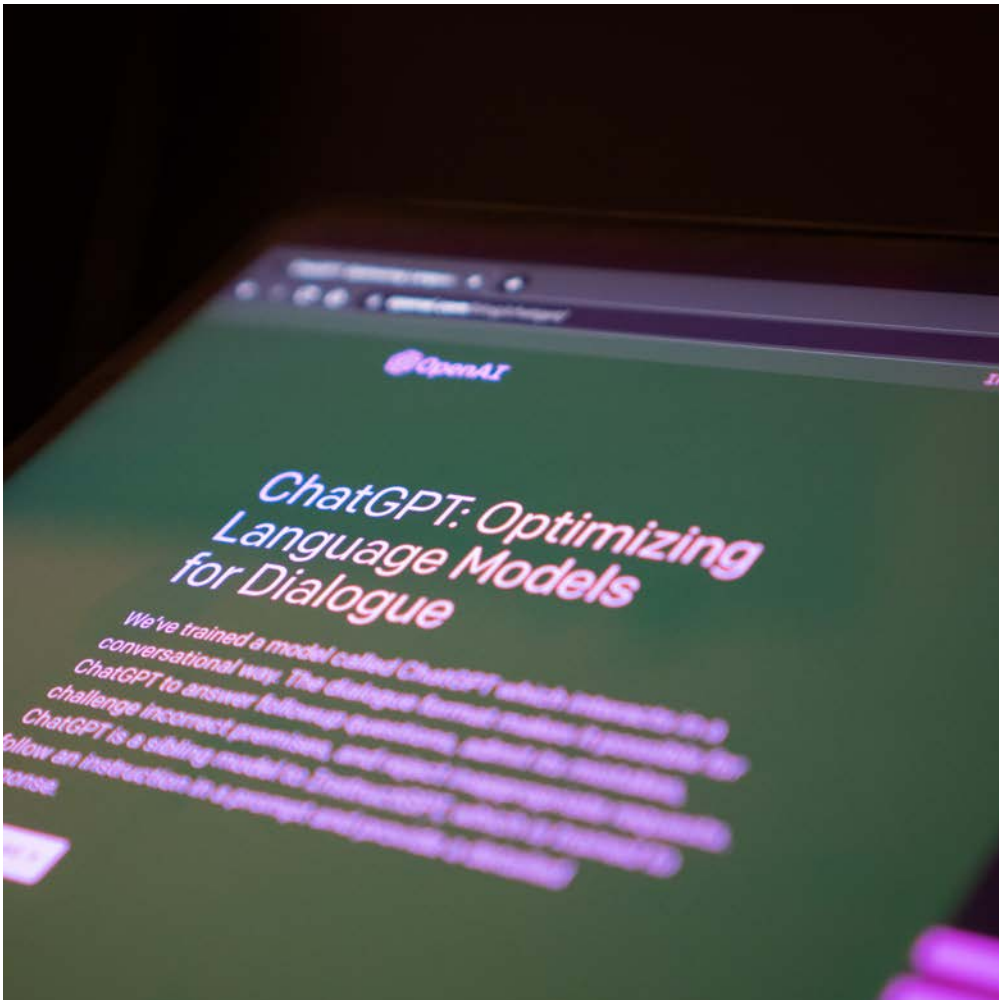
01

法律问题概览

C G

当狂飙的“ChatGPT”遇上法律的缰绳

——速览ChatGPT六大法律问题



ARTICLE BY 周洋 徐颖蕾

2022年11月底，美国人工智能公司OpenAI推出人工智能聊天机器人ChatGPT。此后，ChatGPT在全球掀起一股热潮，国内外互联网巨头纷纷入局。微软先后推出由ChatGPT支持的搜索引擎Bing和浏览器Edge，并在Office办公软件中植入人工智能助手Microsoft 365 Copilot；谷歌发布了其AI聊天机器人“巴德”(“Bard”)，并宣布将生成式AI的功能导入Workspace；2023年8月31日，百度发布的首个“中国版ChatGPT”产品 “文心一言”开放公众使用；9月13日，阿里版的ChatGPT产品“通义千问”向公众开放；9月21日，360宣布其人工智能产品“360智脑”全面接入360“全家桶”并向公众提供服务；京东于今年7月发布产业版ChatGPT“ChatJD”；网易有道于今年7月推出教育领域类ChatGPT产品“子曰”，并已于11月正式通过相关备案；其他知名科技企业的类ChatGPT产品也陆续落地。

那么，ChatGPT[1]的应用会面临哪些法律问题，其背后存在哪些法律风险？本文将从内容管理、电信资质、数据安全和隐私保护、知识产权、不正当竞争等维度对ChatGPT应用中带来的法律问题进行分析。

1.本文为探讨之目的，这里的ChatGPT泛指人工智能生成内容式的技术和产品，不限于OPEN AI公司的产品及其中文版。

/PART 001

如何对ChatGPT生成的内容进行管理?

ChatGPT提供的回答是以大规模的训练数据为基础的。ChatGPT从人们为它投喂的大量数据中学习并生成内容，而训练数据本身的错误、偏见、立场、意识形态和价值观都可能反映在ChatGPT生成的内容中。除了不当的训练数据，算法设计者主观认知偏见、算法设计过程中的技术漏洞，例如算法缺乏信息甄别和过滤机制，都可能影响ChatGPT生成内容的可靠性、正当性。因此，ChatGPT有可能输出不准确的信息，甚至可能输出违法违规或不当信息。一方面，法律、医疗等专业领域的错误答复可能使用户做出错误的判断和决策，从而危害人身和财产安全。另一方面，违法违规或不当信息的传播则可能对公共秩序带来严重后果。因此，如何对人工智能生成的内容进行管理，防止违法违规或不当信息的传播，是ChatGPT在应用中必须首先关注的问题。

根据《网络信息内容生态治理规定》，网络信息内容生产者不得制作、复制、发布含有反对宪法确定的基本原则、危害国家安全等内容的违法信息;且网络信息内容生产者应当采取措施，防范和抵制制作、复制、发布含有低俗、庸俗、媚俗、煽动人群歧视、地域歧视等内容的不良信息；同时，网络信息内容服务平台应当履行信息内容管理主体责任，建立网络信息内容生态治理机制，健全用户注册、账号管理、信息发布审核、跟帖评论审核、版面页面生态管理、实时巡查、应急处置和网络谣言、黑色产业链信息处置等制度。对于ChatGPT而言，其兼具内容生产者和内容服务平台的双重角色，因此，在内容管理措施上也需要同时考虑内容生产和平台管理两方面。

《互联网信息服务深度合成管理规定》就ChatGPT类深度合成服务的提供者如何加强内容管理提出了具体要求，包括：(1)"采取技术或者人工方式对深度合成服务使用者的输入数据和合成结果进行审核"；(2)"建立健全用于识别违法和不良信息的特征库，完善入库标准、规则和程序，记录并留存相关网

络日志"；(3)"发现违法和不良信息的，应当依法采取处置措施，保存有关记录，及时向网信部门和有关主管部门报告；对相关深度合成服务使用者依法依约采取警示、限制功能、暂停服务、关闭账号等处置措施"；(4)"建立健全辟谣机制，发现利用深度合成服务制作、复制、发布、传播虚假信息的，应当及时采取辟谣措施，保存有关记录，并向网信部门和有关主管部门报告"。

此外，如ChatGPT类产品涉及互联网新闻、网络出版、网络直播、网络视听节目、网络文化活动等服务的，还需遵守《互联网新闻信息服务管理规定》《网络出版服务管理规定》《互联网直播服务管理规定》《互联网视听节目服务管理规定》《互联网文化管理暂行规定》等互联网内容服务相关监管规定。

/PART 002

提供ChatGPT服务是否需要电信资质?

根据《中华人民共和国电信条例》，经营电信业务，需依法取得电信业务经营许可证。电信业务分类的具体划分由《电信业务分类目录》列出。此外，根据《互联网信息服务管理办法》，互联网信息服务分为经营性和非经营性两类。其中，"经营性互联网信息服务，是指通过互联网向上网用户有偿提供信息或者网页制作等服务活动"；而"非经营性互联网信息服务，是指通过互联网向上网用户无偿提供具有公开性、共享性信息的服务活动"。另外，"国家对经营性互联网信息服务实行许可制度；对非经营性互联网信息服务实行备案制度。"

ChatGPT通过互联网向用户提供信息，属于互联网信息服务。而对于"经营性"和"非经营性"的判断，不宜简单以服务是否收费来判断有偿或是无偿。实践中，ChatGPT类产品不论是否收费，都具备经营属性，与科研、公益等非经营性活动有明显区分。因此，监管实践中，关于判断是否属于"经营性互

联网信息服务”，监管机关往往会以服务是否符合《电信业务分类目录》所列业务类别进行判定，从而判断是否需要电信许可。

根据《电信业务分类目录（2015年版）》，“B25信息服务业务”是指通过信息采集、开发、处理和信息平台的建设，通过公用通信网或互联网向用户提供信息服务的业务，主要包括信息发布平台和递送服务、信息搜索查询服务、信息社区平台服务、信息即时交互服务、信息保护和处理服务等。其中，“信息发布平台和递送服务”是指建立信息平台，为其他单位或个人用户发布文本、图片、音视频、应用软件等信息提供平台的服务。平台提供者可根据单位或个人用户需要向用户指定的终端、电子邮箱等递送、分发文本、图片、音视频、应用软件等信息。“信息搜索查询服务”是指通过公用通信网或互联网，采取信息收集与检索、数据组织与存储、分类索引、整理排序等方式，为用户提供网页信息、文本、图片、音视频等信息的检索查询服务。

ChatGPT通过对训练数据和用户输入对话的采集、处理以及平台（ChatGPT与用户的交互界面）的建设，通过互联网向用户提供信息内容，符合“信息服务业务”的范畴。从具体的业务类别看，ChatGPT更接近“信息发布平台和递送服务”，而非“信息搜索查询服务”。ChatGPT提供的内容不是经检索与排序的原始信息，而是基于对用户对话的理解和训练数据的分析、编辑后生成的文本。可以理解为，ChatGPT根据用户的要求通过平台向用户提供信息，且ChatGPT本身也参与了信息的生产过程。因此，ChatGPT可能落入增值电信业务中“信息服务业务”的范围，从而该服务提供方需取得B25类“互联网信息服务”的增值电信业务经营许可。

/PART 003

如何处理ChatGPT带来的数据安全和隐私保护问题？

ChatGPT作为史上用户数增长最快的消费者应用，在短短两个月内即突

破了1亿用户。ChatGPT的提供方OpenAI在其官网公布的隐私政策中提到，其产品会收集用户账户信息、对话内容、社交媒体信息、以及Cookies、日志信息、使用情况、设备信息等个人信息[2]。而用户在与ChatGPT进行对话时，可能会进一步透露自己的财产信息、健康信息等敏感个人信息，甚至商业秘密、机密数据。因此，手握大量敏感数据的ChatGPT一旦出现数据泄露、损毁、丢失等安全问题，则可能产生严重的后果。

除了ChatGPT系统漏洞，ChatGPT自身的工作原理也增加了数据泄露的风险。由于用户输入的信息可能被用于进一步训练ChatGPT，而ChatGPT向其他用户输出内容时就可能包含该用户提供的个人信息、机密数据或重要数据，从而引起数据泄露。2023年3月30日，据媒体报道[3]，近日某公司内部发生数起涉及ChatGPT的数据泄露事件。而数据泄露的根源，均是员工将企业机密信息以提问的方式输入到ChatGPT中，导致相关内容进入ChatGPT的学习数据库，从而可能对外泄露。

如果ChatGPT落地中国，还将存在数据出境的问题。根据OpenAI公布的隐私政策[4]，在用户使用ChatGPT服务时，其个人信息都将传输至OpenAI位于美国的设施和服务器上。因此，用户在使用ChatGPT服务中，其与ChatGPT交互时可能提供的个人信息、商业秘密甚至可能关系国家安全、经济运行、社会稳定、公共健康和安全的重要数据都将发生数据的跨境流动。根据目前的数据跨境监管框架，ChatGPT提供服务中如涉及向境外传输重要数据，或ChatGPT处理或者向境外提供的个人信息达到《数据出境安全评估办法》所规定的门槛，则服务提供者需向网信部门申报数据出境安全评估。

2.https://openai.com/policies/privacy-policy，最后访问日期：2023年3月28日。
3."大规模封亚洲IP、遭意大利禁用、泄露芯片机密...ChatGPT遇滑铁卢？"，https://www.thepaper.cn/newsDetail_forward_22643886，最后访问日期：2023年4月11日。
4.https://openai.com/policies/privacy-policy，最后访问日期：2023年3月28日。

/PART 004

ChatGPT采集第三方数据用于训练是否构成"合理使用"?

ChatGPT能够生成各类文本或文案，但这些都来源于对已有作为训练数据的文本或文案的复制、修改、改编、翻译、汇编等的处理。如果前述文本或文案是他人拥有著作权的作品，那么ChatGPT使用作品的行为是否构成对他人著作权的侵犯？2023年3月，据媒体报道，拥有《纽约邮报》《巴伦周刊》《华尔街日报》等的美国新闻集团正准备向OpenAI、M公司和G公司等公司提起诉讼，要求赔偿其内容在ChatGPT、Bard等AI工具中被用来使用的费用。

根据《中华人民共和国著作权法》（以下简称**"《著作权法》"**），使用他人作品应经著作权人许可，并支付报酬，除非符合法律规定的合理使用的情形。《著作权法》第二十四条对于合理使用情形的规定采用了封闭式的列举，而ChatGPT对于训练数据中作品的使用难以符合该法所规定的"个人使用"（为个人学习、研究或者欣赏，使用他人已经发表的作品）、"适当引用"（为介绍、评论某一作品或者说明某一问题，在作品中适当引用他人已经发表的作品）、"科学研究"（为学校课堂教学或者科学研究，翻译、改编、汇编、播放或者少量复制已经发表的作品，供教学或者科研人员使用，但不得出版发行）等合理使用情形。因此，我们理解，ChatGPT使用作品在我国依然需要相应知识产权授权。对于ChatGPT的用户而言，如果直接使用了ChatGPT生成的侵犯他人著作权的内容，也可能面临著作权侵权风险。

目前，已经有国家和地区开始探索将人工智能使用作品的情形作为著作权侵权的例外。例如，日本在其著作权法中将计算机在必要的限度内使用作品纳入合理使用的范畴；欧盟则通过《单一数字市场版权指令》，设置"文本与数据挖掘"的版权例外规则。但我国《著作权法》目前未对人工智能使用作品是否构成合理使用的问题进行回应。

/PART 005

ChatGPT生成内容是否构成作品？谁享有该作品的著作权？

1. 人工智能生成内容的可版权性

根据我国《著作权法》第三条，受著作权法保护的作品，“是指文学、艺术和科学领域内具有独创性并能以一定形式表现的智力成果”。因此，要成为著作权法所保护的作品，不仅需要具有独创性，还需是智力成果。虽然《著作权法》没有明确作品必须为人的智力成果，但通常认为，受著作权法保护的作品必需由人类创造。因此，我国司法实践中往往不承认人工智能生成内容的可版权性。在2019年全国首例人工智能生成内容著作权纠纷案（以下简称**“F案”**）中[5]，法院认为，自然人创作完成应是著作权法上作品的必要条件。人工智能软件利用输入的关键词与算法、规则和模板结合形成的文字内容，某种意义上讲可认定是人工智能软件“创作”了该内容。但即使人工智能软件“创作”的文字内容具有独创性，也不属于著作权法意义上的作品，不能认定人工智能软件是其作者并享有著作权法规定的相关权利。

不过，关于人工智能生成内容是否构成作品，我国司法实践也在进行探索。在上述F案中，虽然法院认为计算机软件智能生成的文字内容不构成作品，但并不意味其进入公有领域，可以被公众自由使用。计算机软件智能生成的文字内容既凝结了软件研发者的投入，也凝结了软件使用者的投入，具备传播价值。[6]而在2020年T公司与X公司侵害著作权纠纷、不正当竞争纠纷一案（以下简称**“D案”**）中，法院对人工智能生成内容的可版权性进行了探索。在该案中，法院认为，软件自动生成文章的过程虽然没有人的参与，但该软件自动运行的方式体现了原告主创团队人员的选择，也由该软件的特性所决定。因

5.参见（2018）京0491民初2xx号F律所与B公司侵害作品署名权纠纷、侵害作品信息网络传播权纠纷民事判决书。
6.参见（2018）京0491民初2xx号F律所与B公司侵害作品署名权纠纷、侵害作品信息网络传播权纠纷民事判决书。

此，从文章生成过程来分析，该文章的表现形式是由主创团队个性化的安排与选择所决定的，体现了人的智力活动，其表现形式并非唯一，具有一定的独创性，构成作品。[7]

此外，2023年11月27日，北京互联网法院审理的一则案件中，法院对原告使用开源软件Stable Diffusion制作的图片认定受著作权保护。尽管法院在该案中对人工智能生成内容的可著作权性提出了突破性观点，但业界对此存在不同看法，并认为该软件生成的内容应适用所谓"知识共享许可协议(Creative Commons Zero 1.0)"，任何人都可以无需授权使用该生成产品，包括出于商业目的。该案对于人工智能生成内容可版权性的判例价值仍有待进一步分析评判。

2. 人工智能生成内容的版权归属

由此引发的问题是，如果人工智能生成内容构成作品，那么著作权归属谁？人工智能生成作品的著作权归属主要涉及人工智能软件的开发环节与使用环节。如果人工智能软件的开发者与使用者竞合，那么权利归属不存在异议，但当人工智能生成软件的开发者和使用者不同一时，人工智能生成内容的著作权归属便存在一定争议。

在F案判决中，法院认为，软件开发者（所有者）没有根据其需求输入关键词进行检索，该分析报告并未传递软件研发者（所有者）的思想、感情的独创性表达；同理，软件用户仅提交了关键词进行搜索，应用"可视化"功能自动生成的分析报告亦非传递软件用户思想、感情的独创性表达，因此，软件研发者（所有者）和使用者均不应成为该分析报告的作者。[8]

而在D案中，法院认为，涉案文章是原告获授权使用D软件后，在原告的

7.参见(2019)粤0305民初14xxx号T公司诉X公司侵害著作权及不正当竞争纠纷判决书。
8.参见（2018）京0491民初2xx号F律所与B公司侵害作品署名权纠纷、侵害作品信息网络传播权纠纷民事判决书。

主持下，由包含编辑团队、产品团队、技术开发团队在内的主创团队运用D软件完成，因此，认定涉案文章是原告主持创作的法人作品，即著作权归软件使用者所有。

对比两个案例可以看出，人工智能生成作品的著作权归属，很大程度上取决于开发者或使用者的智力活动对于人工智能生成内容的独创性的贡献。就ChatGPT生成的内容而言，用户作为使用者大多是以简单的语言文字进行提问，对于人工智能生成内容的独创性作用较为有限。而ChatGPT生成内容更多依赖于其开发者OpenAI的设计、训练和引导。因此，从对内容独创性的贡献上说，开发者OpenAI似乎更符合ChatGPT内容的著作权人。

不过，根据OpenAI的《服务条款》[9]，在法律允许的范围内，OpenAI将所提供的工具（包括ChatGPT）所产生内容的所有权利转移给用户。用户有责任确保生成的内容不违反法律或OpenAI的服务条款。此外，《服务条款》还明确指出，生成的内容不一定具有唯一性，多个用户可能获得相同或非常相似的内容。因此，尽管OpenAI可能被认为是ChatGPT生成内容的著作权人，鉴于OpenAI主动将其权益转让给用户，故该等情形下，ChatGPT生成内容的著作权应当归属于用户。

/PART 006

ChatGPT抓取第三方数据用于训练是否构成不正当竞争？

ChatGPT训练使用的数据大多来自于互联网上公开的网站、信息资源库、数字图书馆、专业数据库、社交媒体平台等。数据收集过程可能涉及利用爬虫协议等底层技术对数据进行搜索、抓取、分析，再用于训练ChatGPT。如果抓取的数据属于数据主体采用技术措施加密或未公开的内容，ChatGPT

9.参见https://openai.com/policies/terms-of-use

的提供方通过绕开数据主体设置的访问限制（比如网站用户对隐私内容设置为"他人不可见"）或绕开部分网站设置的真人审核（例如验证码方式）获取该等数据，那么不仅爬取行为本身可能存在非法获取计算机信息系统数据、侵犯个人信息或商业秘密的风险，ChatGPT使用该等训练数据向用户输出内容还存在不正当竞争的风险。

根据我国反不正当竞争相关司法实践，关于爬虫技术的使用是否构成不正当竞争，法院往往会根据《反不正当竞争法》第二条的原则性条款，即"经营者在生产经营活动中，违反本法规定，扰乱市场竞争秩序，损害其他经营者或者消费者的合法权益的行为"，综合考虑数据抓取方和被抓取方间是否具有竞争关系、被抓取方是否对抓取的数据享有权益、抓取方的行为是否具有正当性、抓取方对抓取数据的使用是否具有正当性、是否给被抓取方造成相应的危害结果等因素后，判断是否构成不正当竞争。对于竞争关系的认定，在互联网经济领域，法院往往采用广义的理解，认为竞争方式主要表现为通过争夺消费者注意力获取竞争优势，实现经营利益，即使经营者之间不存在直接的竞争关系，经营者也因破坏其他经营者的竞争优势与其产生了竞争关系。[10]

因此，ChatGPT完全有可能被认定为与被抓取数据的数据库、社交媒体存在竞争关系。而如果ChatGPT抓取的数据对于被抓取方而言存在商业利益和竞争优势，ChatGPT的抓取行为存在违反Robots协议或法律声明、违反用户协议、违反行业自律公约等不正当的情形，从而对被抓取方的预期利益、合法市场份额、消费者信任度等造成损害，那么ChatGPT抓取训练数据的行为就可能构成不正当竞争。

另一方面，ChatGPT的爆红也引发了另一个不正当竞争问题。ChatGPT虽然未向大陆用户开放，但国内以"ChatGPT""OpenAI"等字眼作为名称的小

10.参见（2020）浙01民终48xx号苏州朗动网络科技有限公司与Y公司等商业诋毁及不正当竞争纠纷民事判决书。

程序、公众号数量激增。这些产品的图标与ChatGPT类似，有些程序号称是ChatGPT的"国内版"，连通ChatGPT的API接口并提供转接服务，实际对话时答非所问，对话质量和ChatGPT相去甚远，显然属于"山寨货"。这些公众号小程序使用与ChatGPT相同或相似的名称及标志，宣传中刻意突出使用"ChatGPT"，误导用户认为与美国人工智能研究实验室 OpenAI 的 ChatGPT 有特定性关联，使消费者产生混淆误认，可能构成《反不正当竞争法》第六条规定的商业混淆不正当竞争行为。

/PART 007

结语

虽然ChatGPT目前并不完美，但它的出现对人工智能产业将产生深远影响，ChatGPT与图片生成、音视频生成、虚拟数字人等工具以及其他AI、云计算等技术集成，都让人们充满期待。人工智能是"人工"的，ChatGPT的研发、训练、使用都离不开人类的设计、控制和规范，与其说人类与ChatGPT等人工智能是人机关系，不如说实质上依然是人与人的关系、是个体与群体的关系（单个个体与人工智能背后古今中外的群体智慧的关系）。如何构建人工智能伦理规范，引导人工智能积极发展，控制人工智能带来的风险，设定技术发展的边界，是人工智能发展的永恒议题，也是人工智能法律规范健全道路上绕不开的重心。

周洋
合伙人
知识产权部
上海办公室
+86 21 6061 3658
zhouyang@zhonglun.com

CHAPTER

02

知识产权

全景透视生成式人工智能的法律挑战(一):

知识产权挑战与合规

ARTICLE BY 陈际红 吴小旭 李佳笑

2023年7月13日，《生成式人工智能服务管理暂行办法》（以下或称"《暂行办法》"）正式发布，并于8月15日正式施行。生成式人工智能（Generative AI，以下简称"AIGC"）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。伴随着《暂行办法》的落地，新一轮AIGC技术的狂飙将面临中国本土的法律挑战。基于此，《全景透视生成式人工智能的法律挑战》系列文章看将从数据合规、知识产权与监管视角，全景透视AIGC所面临的主要法律问题，并准备了《AIGC合规义务清单》，以期为相关企业提供更全面的实操指引。

区别于既定指令的机械执行，"像人类一样思考"的AIGC实现了从"复制"到"创造"的跨越，对现有创作模式产生了颠覆性的变化。鉴于此，本篇在现有知识产权制度语境下，讨论AIGC生命周期可能主要面临的知识产权问题，并尝试提出解决路径。

/PART 001

风险识别：三维度的AIGC知识产权风险管理框架

在知识产权制度语境下，可通过如下三维度的知识产权风险管理框架进行风险识别。

阶段	模型训练	应用运行	模型优化
主要活动	立项设计，数据采集 数据清洗，数据标注 模型训练，模型验证	AIGC服务使用者输入内容，AIGC服务提供者生成内容	利用应用运行阶段采集的内容开展模型优化
主体	· 版权方 · AIGC开发者	· AIGC开发者 · AIGC服务提供者 · AIGC服务使用者（可能为版权方）	· AIGC开发者 · AIGC服务提供者 · AIGC服务使用者（可能为版权方）
焦点问题	作品合理使用的边界	· 生成内容的知识产权属性及权属 · 生成内容侵权风险及责任承担	· 用户输入内容使用的边界

（一）模型训练阶段

主要活动及典型风险：在算法模型训练阶段，AIGC开发者通过对海量数据（其中可能含受著作权保护的作品）进行数据挖掘和信息理解，从而实现算法模型的训练和调试，AIGC实际上是对现有数据（可能含作品）加工整理后的综合式输出结果。此阶段的焦点问题在于通过机器学习进行模型训练是

否构成对作品的合理使用；如果不属于合理使用，是否需要以及如何获得版权方的授权。

主体：AIGC开发者及版权方。

（二）应用运行阶段

主要活动及典型风险：针对内容输入阶段，此等过程涉及由服务使用者输入可能构成作品的内容。针对内容生成阶段，基于AIGC技术底层逻辑，其内容生成天然携带训练所使用的现有作品的记忆，存在对现有作品著作权的侵权风险，进而涉及关于AIGC生成内容的可版权性和权利归属问题。

主体：AIGC服务使用者（可能为版权方）及服务提供者。其中，服务提供者既包括直接向用户提供服务的AIGC开发者，也包括通过引入第三方AIGC开发者技术能力来提供服务的集成方。

（三）模型优化阶段

主要活动及典型风险：通过利用输入内容进行强化学习并进而进行模型优化是AIGC发展的重要过程，此等输入内容可能包含受《著作权法》保护的作品。此外，ChatGPT落地商业运用过程中，关于用户输入内容的衍生处理并进一步对外输出，也导致了多起商业秘密泄露事件发生。据此，需要关注此等模型优化过程的授权问题以及潜在的商业秘密泄露问题。

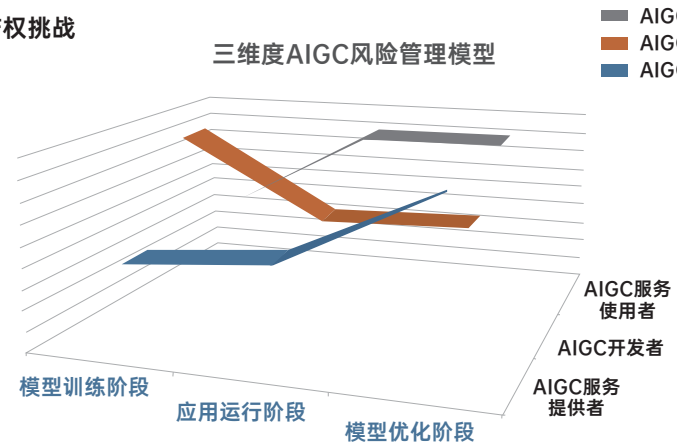
主体：AIGC服务使用者（可能为版权方）及AIGC开发者。在私有化部署及"API-定制化服务"等模式下，集成AIGC技术的服务提供者亦可能构成责任主体。

三维度的AIGC合规方法论

AIGC的知识产权挑战

焦点问题：

- ✓ 作品合理使用边界
- ✓ 生成内容的可版权性及权属
- ✓ 生成内容的侵权风险及责任承担
- ✓ 用户输入内容使用的边界



/PART 002

法律挑战：AIGC对现有知识产权制度的冲击

问题一：在我国制度语境下，使用作品开展模型训练是否构成合理使用？

目前已有多个作者、版权方针对AIGC算法模型训练过程中未经授权的作品使用行为提起诉讼。2022年11月，程序员兼律师Matthew Butterick联合美国Joseph Saveri律师事务所律师，对GitHub Copilot及其背后的微软和OpenAI公司提起诉讼，这是美国第一起关于生成式人工智能的集体诉讼；2023年，美国艺术家对Stability AI在内的三家AIGC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复制其图片用于训练Stable Diffusion的行为提起诉讼。鉴于此，判断利用作品开展AIGC模型训练是否构成合理使用，对于AIGC开发者具有重要意义。

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AIGC开展模型训练构成合理使用任一法定情形的难度较大。具体而言，首先，AIGC的本质是机器学习，且所开发的AIGC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，AIGC作为一种创造性内容创作，并非基于"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AIGC研发一定程度上可以被视为"为科学研究"，但**"少量"**和**"供教学或者科研使用"**的目的限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"，也可认定为合理使用。基于此，《著作权法实施条例》提出了合理使用的"三步检验判断标准"[1]，即应当同时符合"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"。然而，考虑到"合理使用"的认定将会对版权保护产生重要的影响，目前实践中普遍倾向于从严适用"三步检验判断标准"。具体到AIGC场景，由于AIGC生成内容的潜在利用方式多样、利用价值极高，故较难认定"不影响原作品的正常利用"，也很难论证"没有不合理侵害权利人理应享有的合法权益"，而AIGC生成内容又与模型训练密不可分，故目前在我国《著作权法》等制度框架下，论证此等模型训练构成合理使用的难度较大。

问题二：AIGC生成内容是否具有可版权性？

根据《著作权法》规定[2]，针对AIGC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"。AIGC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成为了判断可版权性的重要标准。

近年来，我国司法实践也不乏基于人工智能生成内容是否构成作品的司法判例，具体见下表。

1.《著作权法实施条例》第21条。
2.《著作权法》第三条。

	案例一	案例二
案件事实	• 基于数据库检索数据，利用软件生成包括图形和文字描述大数据报告	• 利用智能写作软件撰写财经新闻报道
裁判思路	• 分析报告符合文字作品的形式要求，具有一定的独创性。 • 具备独创性并非构成文字作品的充分条件，文字作品应由自然人创作完成。 • 分析报告生成过程未体现软件开发者（所有者）、软件用户思想、感情的独创性表达，不构成著作权法意义上的作品。	• 外在表现符合文字作品的形式要求，具有一定的独创性。 • 数据类型的输入与数据格式的处理、触发条件的设定、文章框架模板的选择和语料的设定、智能校验算法模型的训练等均由主创团队相关人员选择与安排，是智力活动。
判决结果	**自然人未参与创作，未体现自然人独创性表达的内容不构成作品**[3]	**构成作品**[4]

在国际上，2023年2月，美国版权局拒绝了含有AIGC生成图片的漫画《黎明的查莉娅》（Zarya of the Dawn）的版权登记申请，认为尽管文本提示影响了人工智能生成内容的方向，但该生成过程缺乏可预测性，不受申请人控制，因而人工智能并非单纯的编辑工具，故申请人可基于文本的作者身份及其对文字、视觉元素的选择、协调和编排，**就文本与图像构成的整体登记版权，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》（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

3.详见（2018）京0491民初239号、（2019）京73民终2030号民事判决书。
4.详见（2019）粤0305民初14004-14007号民事判决书。

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。[5]

可见，无论在中国还是美国，对于AIGC生成内容的可版权性认定思路基本一致：**AIGC生成内容具备独创性且可充分体现人类的智力活动，是AIGC生成内容成为版权法意义上受保护的客体的前提。**

问题三：AIGC生成内容权属属于谁？

对于AIGC生成内容的权属，法律并未就此进行明确规定。目前AIGC相关方一般通过协议等方式对AIGC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AIGC服务使用者，AIGC服务使用者获得相应的使用授权。例如，Open AI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。"[6]

问题四：AIGC生成内容是否存在著作权侵权风险？

由于AIGC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AIGC生成内容，因此，AIGC生成内容天然地、不可避免地携带了训练作品的记忆或痕迹。AIGC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AIGC生成内容与训练作品在表达上构成"实质性相似"且存在"接触可能性"，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

5.关于美国对于AIGC生成内容的监管实践，详见《他山之石丨美国如何认定AIGC的可版权性？》，https://mp.weixin.qq.com/s/F0gg5GG4Ce4pjfujYb1d2g。
6.https://openai.com/policies/terms-of-use，最后访问时间2023年9月12日。

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。

除此之外，由于AIGC生成内容与训练作品的基因脉络一致性，AIGC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，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。例如，使用相同的作品"元素"，可能存在著作权侵权的风险，在金庸诉江南案[7]中，二审法院认为，金庸小说的"人物群像"可以认定为已经充分描述、足够具体，进而得出该"人物群像"属于著作法保护的"表达"的结论。因此，基于对训练作品进行风格模仿、或使用部分作品元素而生成的新"作品"，亦需关注其中的知识产权侵权风险。

CNN的Rachel Metz使用人工智能平台Stable Diffusion通过提示词"Erin Hanson的风格"创建的图画。[8]

Erin Hanson 在2021年创作的油画《水晶枫树》(crystal maples)。

7.详见 （2016）粤0106民初12068号、（2018）粤73民终3169号民事判决书。
8.https://edition.cnn.com/2022/10/21/tech/artists-ai-images/index.html，最后访问时间2023年9月12日。

问题五：AIGC服务提供者需要对生成内容承担侵权责任吗？

在我国，《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。[9]在作品信息网络传播中，网络服务提供者承担侵权责任的前提也是其"知道或应当知道"侵权行为的存在，《信息网络传播权保护条例》对此作出了明确规定[10]。除收到权利人有效通知外，根据《最高人民法院关于审理侵害信息网络传播权民事纠纷案件适用法律若干问题的规定》[11]，是否对作品进行选择、编辑、修改等是"应知"的重要判断因素。但在AIGC的C端业务场景下，生成内容是基于对AIGC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AIGC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、调参入模等而对AIGC生成内容产生影响，但对于最终AIGC生成内容"选择、编辑、修改"的"输入-输出"这一过程，是由AIGC服务使用者与算法共同完成，AIGC服务提供者本身对此控制较为有限，是否可以据此推定AIGC服务提供者对生成内容侵权"明知"仍有待厘清。

值得注意的是，《暂行办法》第9条明确提出AIGC服务提供者应承担"内容生产者"责任，履行网络信息安全义务。此等内容生产者责任是否意味着AIGC服务提供者未来可能会被认定为直接侵权方，而非作为"网络服务提供者"的平台方，亦值得进一步关注。

国际上，美国《数字千年版权法案》（Digital Millennium Copyright Act of 1998）对于网络服务提供者在版权侵权的责任承担限制在"避风港原则"的四种情形范围内。[12]但AIGC生成内容动摇了前述规则适用的前提，将AIGC参与

9.《民法典》第一千一百九十五条。
10.《信息网络传播权保护条例》第二十条、第二十一条、第二十二条、第二十三条。
11.《最高人民法院关于审理侵害信息网络传播权民事纠纷案件适用法律若干问题的规定》第九条。
12.《《数字千年版权法》的现代化路径呼之欲出？》微信公众号：腾讯研究院（ID：cyberlawrc），作者：朱开鑫（腾讯研究院博士后），原文发表于《电子知识产权》2020年第5期。

的内容生成完全剥离为"其他信息内容的提供"或"非内容的网络服务"存在现实困难。例如相较于传统搜索引擎，加载了AIGC的新搜索引擎对于搜索内容整合所呈现的答案显然已经超出了"信息中介"的范围。在234 F.R.D. 674 (2006)一案中，争议焦点即在于内容平台算法推荐服务是否适用《通信正派法案》的责任限制规则，目前美国最高法院已经将案件发回重审，而该案件的最终走向必然也将对该规则的适用产生较大影响。2023年5月16日，OpenAI首席执行官在听证会上也明确表明应当建立针对AIGC新的恰当的监管框架，要求AIGC服务提供者对生成内容承担责任。

/PART 003
解决路径：现有制度框架下的方案设想

AIGC对于内容创作模式的更新，本质上是技术革新对于知识产权既有利益平衡制度的挑战。在我国现有知识产权制度框架下，开展AIGC模型训练以及生成内容的后续利用均会面临诸多法律困境。尽管如此，基于现有制度框架的一些方案设想，或可为相关方提供有益参考。

第一，创设AIGC训练作品的前置管理工具。可参考著作权集体管理制度，由监管部门设立统一的登记机构或成立管理组织，允许版权方自行决定是否将其作品用于AIGC训练，保证版权方对于其版权的控制，当然这一定程度上也会增加AIGC获取训练数据的难度；或者参考开源共享模式，例如，使用知识产权共享协议（Creative Commons license，以下简称"**CC协议**"）建立相关社区以提供训练数据的共享平台。CC协议以简单、标准化的方式赋予作品版权许可，使得该作品能够复制、分发、修改、融合和再创作。[13]版权方可以自主选择对其作品权利保留的范围并进行公开，除保留内容外，在符合协议

13.https://creativecommons.org/about/cclicenses/，最后访问时间2023年9月12日。

约定条件下，其他主体即可以自由地复制、传播等使用相关作品，而无需另行告知版权或获得授权。

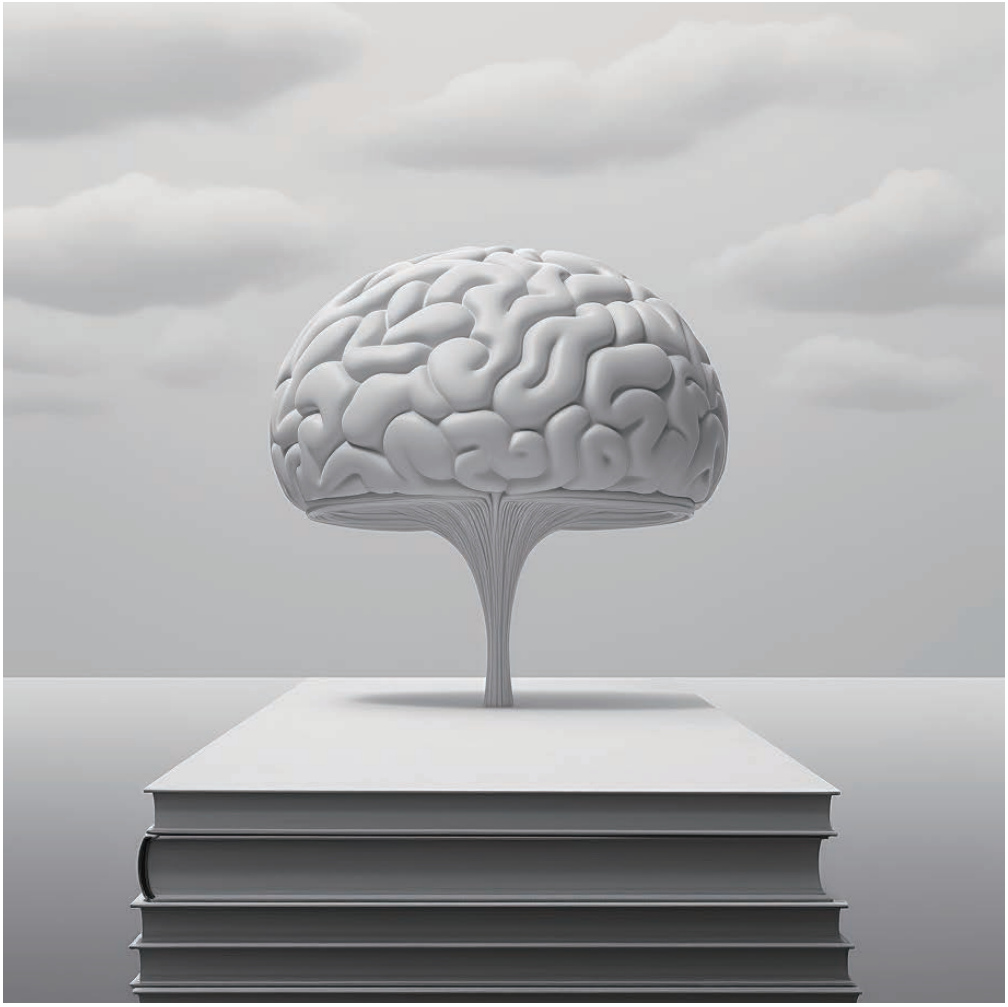
第二，标注 + 退出机制。考虑到前置授权许可的成本问题，也可以针对AIGC生成内容，对涉及作品的使用情况作出标注和说明，并允许作者"选择退出"，以增加作者对于其作品使用的感知和控制。但AIGC训练数据较为庞杂，且生成内容往往并非"直接引用"而可能具有一定的创造性，开展此等标注可能面临技术障碍。另外，对于作者已经提出明确拒绝的作品退出AIGC训练，已经成为行业实践的一般做法。

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

吴小旭
非权益合伙人
知识产权部
北京办公室
+86 10 5087 2938
wuxiaoxu@zhonglun.com

以全球主流AIGC产品用户协议为例

梳理AIGC生成内容的权利归属与使用限制



ARTICLE BY 王飞

AIGC（人工智能生成内容，全称"AI Generated Content"）产品生成的内容若能够满足作品"独"和"创"的构成要件，应当认定构成"作品"，受到《著作权法》的保护。AIGC生成内容的过程离不开软件开发者的参与，软件开发者前期进行的开发与训练赋予了AIGC软件生成内容以独创性；用户作为启动程序和输入内容的人，也参与到对算法的调整中，使得生成内容能够体现其一定的个性选择与判断。由此，从基本原理角度来看，AIGC生成内容的著作权应由软件开发者和用户共同享有。但正如笔者在前述文章所指出的，约定优先原则仍可适用。前述著作权共有是基于著作权法原理进行的权利首次分配，而全球主流AIGC产品一般通过用户协议对AIGC生成内容的权利进行二次利益分配，以适应不同AIGC产品商业模式、底层模型的工作原理、生成内容侵权风险差异和软件开发者对生成内容的使用需求。本文将以全球主流AIGC产品用户协议为研究样本，梳理出不同AIGC产品对于AIGC生成内容权利归属约定的类型以及对于用户使用AIGC生成内容作出的限制，从而探究其背后的法律原理和商业逻辑，为未来AIGC产品在权利归属、责任承担、风险规避等层面提供借鉴。

/PART 001

全球主流AIGC产品生成内容权属约定类型

笔者认为，著作权作为一种私权利，在不违背法律法规强制性规定的情况下，当事人可以对创作生成的内容通过协议的方式约定权利归属。但著作权共有必然会导致权利和责任分配不便、未来权益行使困难等问题，由此，主流AIGC产品均未采取著作权共有的模式。笔者以全球主流AIGC产品用户协议为例进行梳理发现，AIGC产品生产内容的权利归属模式可分为以下五种类型。

1、权利全部归属于用户

将AIGC产品生成内容的全部权益转让或直接约定为用户所有是全球主流AIGC产品的常见做法之一。例如，OpenAI的用户协议规定："在用户遵守使用条款的前提下，OpenAI向用户转让其对输出享有的所有权利、所有权和利益；由其他用户要求并为其产生的回应不被视为您的内容。"该条款还进一步保证了不同用户生成内容之间的独立性，以避免单个用户对相同输出结果主张独占性权利。Anlatan旗下创作辅助类AIGC产品Noval AI在用户协议中明确："用户保留对其内容的所有权利和所有权；我们不要求对用户内容享有任何所有权，除非用户同意并经Anlatan特别同意，将用户的所有权转让给Anlatan。"某开源平台旗下一款能够自动补全代码的AIGC产品规定："（本产品）返回给您的代码、函数和其他输出被称为'建议'。（本平台）不对建议主张任何权利，您保留对您的代码（包括您在代码中包含的建议）的所有权和责任。"

2、权利归属于用户，软件开发者取得授权

为保证软件开发者商业利益，即使约定AIGC产品生成内容的权益归属于用户，部分AIGC产品也会特别要求取得用户授权，以对用户内容进行展示、储存、发布、复制、使用和修改等，进而实现软件开发者利用用户内容进行盈

利、宣传和软件升级等目的。例如，Canva可画约定："你（用户）向Canva授予一项免费的、可再许可的授权，允许Canva在为你提供服务的必要范围内展示、储存、发布、复制和使用你的用户内容。"Notion AI约定："客户特此授予Notion全球范围内的、非独家的、不可撤销的、免版税的、全额支付的、可分许可的（给Notion的第三方服务供应商）许可，以托管、存储、转让、展示、执行、复制、修改、创造衍生作品，以及分发与向客户提供服务有关的客户数据。"Jasper AI约定："通过提交、张贴、展示、提供或以其他方式在服务上或通过服务提供任何客户内容，您明确授予，并且您声明和保证您拥有所有必要的权利，并授予Jasper免版税、可转授权、可转让、永久、不可撤销、非独家的全球许可，以使用、复制、修改、出版、列出有关信息、编辑、翻译、分发、联合、公开表演、公开展示，以及制作所有这些客户内容和您的名字的衍生品、您的客户内容中包含的全部或部分您的姓名、声音和/或肖像，并以任何形式、媒体或技术（无论是现在已知的还是以后开发的）用于与服务和Jasper（及其继承人和附属机构）的有关业务，包括但不限于与修改、改进和提高人工智能模型有关，以及以任何媒体格式和通过任何媒体渠道推广和重新分发的部分或全部服务（及其衍生作品）。"ChatGPT则是单独规定了OpenAI可以将用户内容用以改进和提升服务，也就是使用用户输入内容来进行机器学习。若用户不同意自己的输入内容被用于上述使用，则可以通过邮件联系OpenAI表示拒绝，但可能会影响用户在特定场景下对ChatGPT的使用。

3、权利归属于软件开发者，用户取得授权

较少数AIGC产品在用户协议中约定由软件开发者享有用户生成内容的权益。例如，一款漫画分镜类AIGC产品Storyboard That在其商用限制条款中规定："免费用户不能使用其Storyboard That创作来盈利；购买了至少一年高级账户的用户可以对创作内容进行发布，但是只能在出版物中引用Storyboard

That，同时需要告知出版商，其并不拥有插图的权利。"从中可见，用户利用Storyboard That生成的内容权利归属于Storyboard That，而用户只拥有使用的权利。

4、以是否付费确定权利归属于用户

部分AIGC产品以用户是否进行付费使用产品，将用户区分为"免费/普通用户"和"付费/会员用户"，进而以此为依据确定权利归属。例如，会员用户可以获得生成内容的全部权益，而普通用户生成内容只能获得使用授权。

例如，AI绘画软件Midjourney约定付费会员用户拥有其用服务创建的所有资产，但不包括对他人图像的放大；非付费用户不享有其所创作的资产，Midjourney在知识共享许可协议4.0版（Creative Commons license 4.0，简称**"CC 4.0协议"**）下给予其授权。某国内AI绘图软件约定："在ERNIE-ViLG AI作画大模型体验专区生成的内容均带有水印，用户在付费后使用API调用时生成的图像将全部去除水印。除法律法规和本协议另有相反规定，无水印图像的知识产权及其上的相关权益（包括但不限于知识产权等）将永久归用户所有。"但值得注意的是，相关协议并未直接明确未付费情形下生成内容的权属，不过从协议约定的方式来看，我们认为，软件开发者应是以默认归属于软件开发者、用户付费后转让权益给用户的逻辑进行约定的。

5、流入"公有领域"

较为特殊的是文生图类AIGC产品Stable Diffusion Online，其将生成内容以适用知识共享许可协议（Creative Commons Zero 1.0，简称"**CC0 1.0通用协议**"）的方式流入"公有领域"，任何人都可以通过复制、修改、发行等方式利用生成内容，包括商业目的，而无需获得事前授权。

/PART 002

权属约定背后的商业逻辑

1、生成内容侵权风险的高低

鉴于不同的AIGC产品商业逻辑、底层数据和商业盈利模式不同，进而呈现出对于AIGC生成内容权利归属的不同约定。基于笔者对主流AIGC产品商业模型研究，约定权利归属不同的主要原因在于生成内容侵权风险不同。而基于权责统一的基本原理，将生成内容权利归属于用户对于发展初期的AIGC产品而言似乎是最安全的发展模式之一，特别是考虑到生成内容可能有较高的侵权风险。例如，Stable Diffusion甚至选择放弃生成内容相关权益，主要原因可能来源于其生成内容相对较高的侵权风险。从Stable Diffusion的工作原理来看，在模型训练阶段，其需将数据库中的版权图片作为输入对象，对其添加"噪点"并编码，并与描述性文本进行交互，形成"潜在表现形式"。在内容输出阶段，模型会对"潜在表现形式"进行"去噪"，最终得到新的图像内容。暂且不论其使用版权方权利作品进行模型训练是否构成侵权，"去噪"与解码后生成的内容本身即可能侵犯原作品的复制权和改编权。在马里兰大学和纽约大学的联合研究中，研究人员发现利用Stable Diffusion生成的内容与底层模型训练所使用的数据集作品相似度超过50%的可能性达到了1.88%。由此，对于Stable Diffusion而言，放弃生成内容权益本身可能是其规避侵权纠纷的"权宜之计"。即使是与原作品相似可能性较低的ChatGPT亦同样选择将生成内容权利归属于用户，以尽可能减少在发展初期可能面临的侵权纠纷。

相反，Storyboard That是为数不多采取生成内容归属于软件开发者的AIGC产品之一，究其原因，主要是由于Storyboard That的生成内容全部由其图库中超过2500万个人物造型、物品和场景等原始图像组成。生成内容中的全部元素均为独立的美术作品且由Storyboard That享有权利，只不过排列组合方式不同。因此，Storyboard That对于生成内容可能导致的侵权风险具有可预

见性，从而约定生成内容权利归属于自身。

2、AIGC产品商业利益需求和保护

除避免陷入侵权纠纷困局外，AIGC产品软件开发者对于生成内容的商业需求也影响着对于权利分配的约定。鉴于生成内容对于AIGC产品的宣传、盈利、进一步机器训练、衍生创作等起着重要作用，由此，若约定生成内容权利归属用户，往往AIGC产品软件开发者会再设置授权条款，将用户享有权益的生成内容再授予软件开发者一项使用权利，以保证软件开发者使用生成内容不受限制。而与此同时，为避免用户将生成内容用于其他商业化开发或竞品相关而导致自身产品的权益受损，软件开发者还会通过使用限制条款对用户权益进行限制。例如，Canva可画的用户协议就禁止用户进行下述使用行为：(1)为分发和/或者销售给第三方，在网站、文档或其他模版中使用付费内容；(2)将任何内容（免费字体除外）用作商标、设计标记、品牌名称、字号、企业名称、服务标记或徽标的一部分；(3)不得在网站或者其他地方单独使用或展示内容，以促进或涉及产品的销售、许可或其他分销等。

3、最大可能降低AIGC产品责任

除通过约定生成内容权利归属方式规避侵权纠纷外，为避免AIGC产品在前景不明朗的情势下陷入海量侵权纠纷，进一步通过赔偿条款、免责声明条款和责任限制条款也成为主流AIGC产品的选项。无一例外，无论生成内容权属怎样约定，对于用户使用AIGC服务或违反用户协议而引起的责任，主流AIGC产品均约定由用户承担。即使是约定"软件开发者享有生成内容权益"的Storyboard That也如此规定："对于因用户使用服务或用户违反本协议的任何规定而引起的任何及所有的索赔、诉讼、程序、控告、责任、损失、损害、费用、支出和律师费，用户应维护、赔偿并使Clever Prototypes免受损害。"

不仅如此，除从正面约定用户承担因使用服务或违反用户协议而引起的一切责任外，AIGC产品还通过免责声明条款拒绝对AIGC产品服务作出质量、适用性、无侵权等保证，从而排除相关责任。例如，OpenAI协议约定："本服务按'如其所示'提供，除非在法律禁止的范围内，我们不对服务作出任何保证，包括但不限于对适销性、特定目的适用性、质量、非侵权保证，以及由任何交易过程或贸易惯例产生的保证；我们不保证服务将是不间断的、准确的或没有错误的，或任何内容将是安全的或不丢失或改变的。"

为进一步排除AIGC产品开发者承担侵权责任的可能性，对于非因用户使用服务或违反用户协议而引发的且未落入免责声明条款范畴的损失，主流AIGC产品用户协议还通过约定责任限制条款，以将AIGC产品开发者的风险控制在可承受范围内。例如，全球主流AIGC产品用户协议中的责任限制条款均免除了AIGC产品开发者对任何间接性、偶然性、特殊性、后果性、惩戒性损害（例如商誉损失、停工、计算机故障等）的责任，无论产品使用者是否已被告知可能发生这种损害。除此之外，部分AIGC产品的用户协议还进一步设定赔偿数额上限。例如，OpenAI协议约定："我们的总责任不应超过您在责任产生前12个月内为引起索赔的服务所支付的金额或100美元，以较高者为准。"

/PART 003
结语

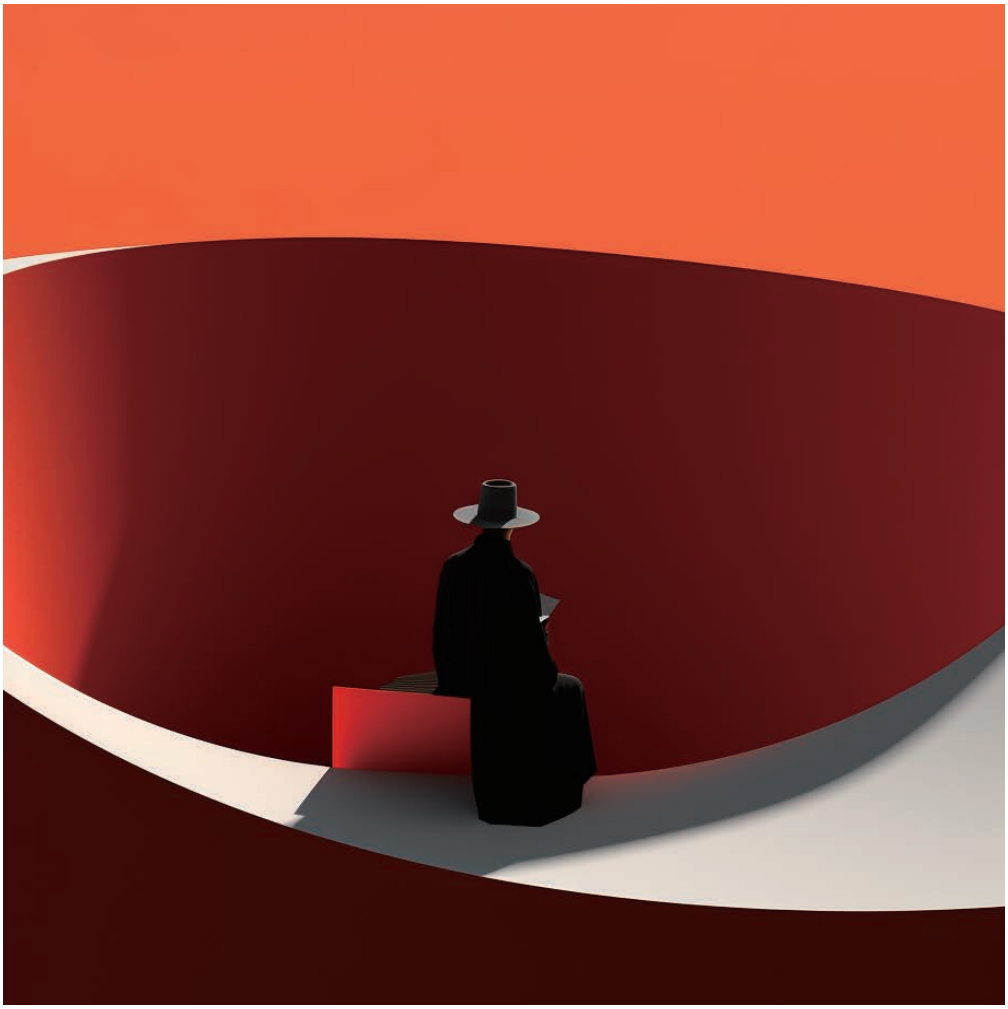
基于对全球主流AIGC产品生成内容权利归属约定的梳理可以看出，将生成内容权利归属于用户、AIGC产品取得授权是更为安全的发展模式之一。当然，若是产品调用的底层数据权属明晰且对于侵权风险具有可预知性，权属约定为软件开发者亦是可考虑的思路。除此以外，通过使用限制、赔偿条款、免责声明条款和责任限制条款等约定尽可能避免AIGC产品陷入侵权纠纷尤为重

要。未来AIGC产品需结合自身商业模式、内容生成的底层数据逻辑、商业盈利模式等，选择最为适宜的约定方式。

（洪妍对本文亦有贡献）

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

人工智能企业知识产权管理实践探讨



ARTICLE BY 王红燕

根据相关统计，目前人工智能产业正从发展期向成熟期过渡，步入了稳步增长阶段，2022年人工智能核心产业规模预计达到2476亿元规模，带动相关产业规模9396亿元，到2026年预计核心产业规模将超过6000亿元。其中，核心产品包括计算机视觉、智能语音、对话式AI、机器学习（含自动驾驶）、知识图谱、自然语言处理、AI芯片等。随着人工智能产业的发展，也打开了新一轮的城市和区域竞争变局。根据中国新一代人工智能发展战略研究院2018-2021年针对区域人工智能科技产业竞争力评价指数的追踪研究表明，2021年长三角总评分首次超过京津冀位列第一。人工智能企业主要集中在应用层，京津冀以及长三角地区基础层、技术层企业占比高于珠三角及川渝地区。[1]

知识产权是重要的无形资产，根据智力资本商业银行公司OCEAN TOMO调查显示，1975年标准普尔500企业的市值组成中，83%是有形资产（工厂、机器、房产等），而到了2015年，企业市值的84%已由无形资产决定，主要是知识产权。[2]在人工智能被认为是未来国际竞争制高点的格局下，人工智能企业作为创新主体，必须提高知识产权的管理水平，这样才有可能保证企业的竞争力和可持续发展。

笔者结合在人工智能企业的工作经历和案例，将从人工智能企业知识产权管理策略和重点、人工智能企业知识产权管理制度、人工智能企业知识产权保护及风险管理三方面展开讨论。

1.艾瑞咨询，2021年中国人工智能产业研究报告，www.iresearch.com.cn
2.国家知识产权局知识产权保护司，企业知识产权保护指南

/PART 001

人工智能企业知识产权管理策略和重点

企业要做好知识产权管理，首先要明确知识产权管理的目标，而就这一点来说，人工智能企业与其它产业并无不同，其根本目标都是在与企业的总体经营目标相匹配的前提下，提升企业的商业竞争力。在知识产权管理目标确定后，就需要进一步制定知识产权管理策略，明确知识产权管理重点。

通常来说，在制定知识产权管理策略时，需要考虑企业的主营业务类型、企业规模、市场地位、财务预算等多种因素，并且在不同发展阶段，适时调整策略以适应企业的发展。例如，对于初创型人工智能企业，应该更关注核心知识产权的保护，尤其是对占据细分市场具有强竞争力的自主创新技术，并且相比知识产权的数量和广泛布局，更需要关注知识产权的质量。此外，在知识产权保护区域的选择上，要优先选择法律制度健全且产品销量较大的主要市场，在有限的财务预算内做好最优的资源配置。

随着技术的发展，知识产权保护对象的外延也在不断扩展，整体上包括专利、商业秘密、版权、商标、域名、数据等类型，不同类型的知识产权具有不同的特点。例如，专利权是国家依法在一定时期内授予专利权人或者其权利继受者独占使用其发明创造的权利，其保护范围较广，可保护产品（包括形状、构造或其结合）、方法或者其改进有关的新技术方案，但授权需要经过国家知识产权局审查，且满足新颖性、创造性和实用性等法定要求，流程较为复杂；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，无保护期限限制，但不是绝对垄断的权利，权利证明和侵权取证难度都较大；软件著作权保护计算机程序及其有关文档，自作品完成之日起即自动享有，但保护不延及开发软件所用的思想、处理过程、操作方法或者数学概念等。企业在进行知识产权管理时应熟悉不同类别知识产权的特点，并据此把握知识产权管理的侧重方向。人工智能企业发展的关

键要素包括算法、数据、算力，因此，在知识产权管理方向上，不同于传统行业，人工智能企业将更侧重于人工智能算法与芯片的商业秘密和专利保护、计算机软件的著作权保护、数据合规和安全等方面的知识产权管理。

/PART 002

人工智能企业知识产权管理制度

从国家层面来讲，知识产权法律制度的主要目的是在保护公众利益的同时，最大限度地激励人们进行发明创造、智力创作或者合法经营。自然地，落实到企业层面，企业的知识产权制度同样要起到激励创新的作用。因此，创新激励制度是人工智能企业必不可少的知识产权管理制度之一，包括专利奖励制度、标准参与及制定奖励制度、创新项目申报奖励制度等。

人工智能企业的创新源泉通常来自技术人员，但不同于八九十年代，目前技术人员的流动性和跳槽频率明显提高，导致职务发明纠纷频发，这一点可以在科创板上市企业的问询中体现。自2019年6月13日科创板开板以来，核心技术人员在拟上市企业中所研发形成的核心专利是否与原单位或者高校存在权属争议，一直是上市委重点关注的问题。例如，人工智能企业虹软科技，在上市申报过程中被上市委要求回复问题11："关于知识产权请发行人：……（2）说明实际控制人、董事、高级管理人员、核心技术人员是否存在违反与曾任职单位之间的竞业禁止协议或保密协议的情况……（5）D某与Z某发明专利数占公司所有发明专利比例高于10%，两人已于2005年从公司离职，以上两人离职的原因，是否与发行人在专利权属方面存在纠纷。"由此可见，职务发明管理制度同样是人工智能企业必不可少的知识产权管理制度之一。

无论是初创型企业还是大型企业，在具体到知识产权保护的操作层面时，都需要制定相应的知识产权保护评审制度，对内部的软件、硬件技术进行筛选和识别，判断哪些技术需要保护，选择什么手段进行保护，保护的级别是什么

等等。例如，可制定专利保护评审制度，通过设立由发明人、技术专家、专利工程师、市场人员等组成的专利评审委员会，定期对技术交底书进行评审，评审维度可包括技术先进性、市场应用范围、可专利性等，由此得出技术交底书中的内容是否适合申请专利、选择哪种专利申请类别（发明、实用新型、外观设计）、专利的等级是什么（核心、外围、预研）、专利申请的地域选择等结论，从而保证申请的专利具有真实价值。同时，如能配合较高的专利撰写水平，则更能保证有较高的专利质量。这样能够切实保护到企业的技术，也能构建技术壁垒。

由于在落实知识产权管理的过程中，始终离不开人的意识和行为，只有在企业内部培养和构建知识产权保护意识，形成良好的知识产权保护文化和氛围，才能更好地推进工作。由此，必须建立完善的知识产权培训制度，对不同部门、不同层级的企业人员进行分类和分级培训。例如，分类培训可以包括：1）对知识产权内部团队进行相关法律法规、经典案例、专利挖掘、检索分析、布局等方面的培训，提升知识产权管理人员的专业能力和水平；2）对市场销售团队就商务拓展中的专利风险、侵权取证保全、专利资产的运用等方面进行培训；3）对人力资源团队就入职离职调查制度、奖励考核制度等方面进行培训；而分级培训可以包括：1）与企业高层管理者就企业所处的竞争环境、企业所面临的知识产权风险、挑战和机遇等方面进行讨论和规划；2）与企业中层管理者进行沟通使其与高层制定的战略目标和策略达成共识，并制定细化的实施方案；3）对企业基层员工进行侧重于技术交底书撰写、专利挖掘等具体操作实务方面的培训；4）对企业新员工进行侧重于知识产权相关制度的培训。

此外，根据企业的不同发展阶段，可逐步完善和细化知识产权管理制度，建立数据合规审查制度、开源软件审核规范、技术进出口审核机制、人工智能伦理安全风险治理等管理规章内容。

/PART 003

人工智能企业知识产权保护和风险管理

如前所述，人工智能企业的知识产权管理更侧重于人工智能算法与芯片的商业秘密和专利保护、计算机软件的著作权保护、数据合规和安全等方面。因篇幅所限，本文将仅针对专利、商业秘密和数据合规三方面展开讨论。

1.专利保护和风险管理

首先，根据企业的不同发展阶段，专利保护工作具有不同的目标和重点。在企业起步阶段，通常专利保护工作在合理布局的前提下以专利资产积累为主，工作重点主要在专利挖掘、专利申请及流程、文档管理等；在企业发展阶段，除继续积累专利资产以外，还需要进一步提高对专利质量的要求，增加风险控制和防范的需求，工作重点将同样增加专利检索与分析，以及专利风险管理等；在成熟阶段，企业将增加对专利资产运营的需求，此时工作重点将增加专利价值评估、维护管理及运营管理等。

其次，企业的专利保护工作还需从不同维度展开，包括1）专利申请数量方面：应当与同行可比竞争对手相比，具有合理的专利申请数量及授权数量；而针对科创板拟上市企业，还需注意满足《科创属性评价指引（试行）》4项常规指标之（3）的要求——形成主营业务收入的发明专利5项以上；2）专利保护内容方面：可根据专利评审结果，优先保护核心技术，然后保护与核心技术相关的外围技术，在企业财务及规划允许的情况下，再增加对前瞻性技术、与竞争公司对抗的相关技术的保护；同时，要注意保障核心专利与主要产品、核心技术、主营业务收入的对应关系；3）专利申请时间方面：在技术开发过程中，应持续做好专利保护工作，确保核心专利的申请时间与核心技术的迭代时间相匹配，并且务必在技术公开之前提交专利申请，避免因市场推广活动或开源共享而导致技术被提前公开最终影响专利的授权；4）专利申请地域方

面：要综合考虑公司的市场规划、拟申请地域的法律和资源环境、竞争对手的布局地域等；5）专利发明人方面：一些企业存在发明人挂名现象，实践中，应尽量避免，以防后期产生法律纠纷。

随着人工智能创新的快速发展，全球人工智能相关专利申请以平均每年28%的速度增长，中国人工智能技术专利同样也保持爆发增长态势。截止2021年9月，中国人工智能领域申请专利共计909401件，授权专利253811件。[3]庞大的专利数据背后，也暗藏诸多专利风险，因此，人工智能企业在进行专利风险管理时，必须关注专利侵权风险，结合FTO分析结果评估侵权风险的高低，并配合规避设计、无效请求、现有技术抗辩、先用权抗辩、合法来源抗辩、寻求许可转让等手段进行风险应对。同时，还需关注专利的使用风险，在合作或委托开发过程中避免出现未对使用权和收益权进行约定或约定不明的情况，且尤其关注涉诉专利的使用风险。

2.商业秘密保护和风险管理

由于目前各国对人工智能专利申请的保护客体的适格性判断上有不同的标准，人工智能技术方案容易被认定为"抽象概念"、"智力活动的规则"而被排除在保护客体之外。并且，人工智能专利因为可视化程度较低，存在举证困难的问题，使得人工智能技术在专利保护和维权上存在一定的难度。而在人工智能领域，数据、算法、计算机程序及其有关文档等非公知信息是主要的技术秘密，因此，人工智能企业在做好专利保护和风险管理的同时，尤其需要关注商业秘密的保护和风险管理。

根据现行《反不正当竞争法》第九条第四款规定，商业秘密需要具备秘密性、价值性以及保密性。在商业秘密侵权纠纷案件中，权利人必须先行明确其商业秘密的秘密点，否则将会因为权利范围不清楚而难以获得法律保护。例

3.国家工业信息安全发展研究中心、工信部电子知识产权中心，《中国人工智能高价值专利及创新驱动力分析》报告

如，人工智能企业在智能软件的开发过程中通常会在开源代码的基础上开发，若针对开源代码部分并没有修改，则法院会要求在除去开源代码的其余有独创性的代码范围内确定秘密点。而在"中国无人驾驶第一案"的侵害商业秘密纠纷案中，虽然原告公司认为被告在原告公司在职及离职期间的下述行为构成对原告公司商业秘密的侵犯：1.违反竞业协议，创立与原告公司有直接竞争关系的公司并挖走部分核心员工；2.离职时未上交存储有原告公司商业秘密的物品；但因为原告公司始终无法证明商业秘密的存在，也无法证明被告有侵害商业秘密的行为，原告公司最终撤回起诉。[4]

因此，在进行商业秘密管理时，应明确哪些属于商业秘密，依据商业秘密的价值、商业秘密的载体形式等具体情况对商业秘密设定不同的密级、保密期限和保密措施。很多人工智能企业会涉及到算法推荐服务，《互联网信息服务算法推荐管理规定》第十六条要求算法推荐服务提供者应当以显著方式告知用户其提供算法推荐服务的情况，并以适当方式公示算法推荐服务的基本原理、目的意图和主要运行机制等。人工智能企业在披露算法规则时应注意避免涉商业秘密的细节，可通过示例、流程图等方式告知用户满足合规要求。例如，国内某头部外卖平台在2021年9月首次公开骑手配送中"预估到达时间"的算法规则，其中并未披露技术细节，但对于基本原理（四层算法模型）、目的意图（为骑手提供充裕送餐时间）、运行机制（在模型预估时间的基础上增加三层保护时间）均予以告知。[5]

实践中，大部分商业秘密侵权纠纷的产生与员工离职有关，在人工智能领域，人才的流动相对于传统行业显得更为频繁，但是公司的技术秘密往往与员工自身的知识和技能交织在一起，有时很难确定两者的界限，商业秘密被披露的风险自然难以避免。因此，除了要求员工签署保密协议之外，往往还需通过与员工签署竞业限制协议的方式，避免员工离职导致的商业秘密泄露或被不正

4."中国无人驾驶第一案"尘埃落定，法庭宣判在即、百度撤诉王劲，http://k.sina.com.cn/art
5.陈际红，陈煜烺，算法推荐新规生效：五大视角厘清算法治理新格局

当使用。

不管是有意泄漏还是无意泄漏，商业秘密信息一旦泄露，就不再是商业秘密，人工智能企业可根据《关于审理侵犯商业秘密民事案件适用法律若干问题的规定》第六条的规定，采取保密措施，包括：（1）签订保密协议或者在合同中约定保密义务；（2）通过章程、培训、规章制度、书面告知等方式，对能够接触、获取商业秘密的员工、前员工、供应商、客户、来访者等提出保密要求；（3）对涉密的厂房、车间等生产经营场所限制来访者或者进行区分管理；（4）以标记、分类、隔离、加密、封存、限制能够接触或者获取的人员范围等方式，对商业秘密及其载体进行区分和管理；（5）对能够接触、获取商业秘密的计算机设备、电子设备、网络设备、存储设备、软件等，采取禁止或者限制使用、访问、存储、复制等措施；（6）要求离职员工登记、返还、清除、销毁其接触或者获取的商业秘密及其载体，继续承担保密义务等。

3.数据保护和风险管理

人工智能的实现，与海量的数据基础密不可分，在人工智能技术的运用过程中，通常会涉及到数据的收集、存储、处理和使用。

在数据收集环节，收集的数据内容可能包括用户姓名、出生日期、证件号码、电话号码、住址、生物识别信息、邮箱、健康信息等个人信息，收集渠道可能包括从第三方购买、自行收集、网络爬虫等多种方式。《民法典》《数据安全法》及《个人信息保护法》均规定自然人的个人信息受法律保护，《网络安全法》规定个人信息控制者应当"遵循合法、正当、必要的原则"，并禁止收集"与其提供的服务无关的个人信息"。此外，根据《刑法》第253条之一规定，侵害他人的个人信息不仅要承担民事责任，还可能承担刑事责任。人工智能企业在数据收集环节，需尤其注意合法获取个人信息。在"中国人脸识别第一案"——郭兵诉杭州野生动物世界有限公司服务合同纠纷中，二审法院认为："人脸识别信息相比其他生物识别信息而言，呈现出敏感度高，采集方式

多样、隐蔽和灵活的特性，不当使用将给公民的人身、财产带来不可预测的风险，应当作出更加严格的规制和保护。经营者只有在消费者充分知情同意的前提下方能收集和使用，且须遵循合法、正当、必要原则。野生动物世界在涉指纹识别的'年卡办理流程'中规定'至年卡中心拍照'，郭兵亦同意在办卡时拍摄照片，但提供照片仅系为了配合指纹年卡的使用，不应视为其已授权同意野生动物世界将照片用于人脸识别。野生动物世界虽自述其并未将收集的照片激活处理为人脸识别信息，但其欲利用收集的照片扩大信息处理范围，超出事前收集目的，违反了正当性原则。同时，鉴于收集照片与人脸识别利用的特定关系，野生动物世界又以短信通知等方式要求郭兵激活人脸识别，表明其存在侵害郭兵面部特征信息之人格利益的可能与危险。"[6]

数据存储贯穿于人工智能技术的整个过程，最大的风险在于因系统漏洞而导致数据的泄漏。例如，在北京T技术有限公司、北京T科技发展有限公司与某网络技术有限公司不正当竞争纠纷案中，二审法院认为，被上诉人某网络技术有限公司经营的新浪微博拥有数亿用户，通过Open API向众多第三方应用软件提供接口，其作为Open API平台提供方，在其认为没有授予上诉人北京T技术有限公司、北京T科技发展有限公司相应权限的情况下，上诉人北京T技术有限公司、北京T科技发展有限公司已然通过Open API接口获取了相应信息，暴露出被上诉人对于Open API权限控制的漏洞，其在Open API接口控制权限的设置、信息通过Open API接口调用的检测以及调用过程的记录等方面存在严重的缺陷。鉴于Open API合作开发模式的巨大潜力以及在互联网大数据时代的积极作用，互联网企业在运用Open API开展合作开发时，不仅应将用户数据信息作为竞争优势加以保护，还应将保护用户数据信息作为企业的社会责任，采取相应的技术措施提升Open API合作模式中相应权限的控制，不断完善Open API合作模式。[7]

6.（2020）浙01民终10940号
7.（2016）京73民终588号

人工智能企业在进行数据相关处理事项时，应注意遵守合法、正当、必要和最小范围原则，获取权利人的充分授权，不过度收集个人信息，在存储和使用的过程中采取必要措施，确保数据安全，防止数据泄漏、被窃取、被篡改或丢失，不非法出售或未经授权向他人提供数据。同时尽量对合法获取的数据进行清洁处理，避免侵犯个人隐私；在收集、处理和使用过程中遵守公开处理原则，确保数据主体的知情权和控制权。另外，如果涉及到业务确需向境外提供数据，且属于《数据出境安全评估办法》规定的特定情形的，应当对数据出境进行安全评估申报和风险自评估。

/PART 004
结语

人工智能技术的发展，对国家、经济和社会的进步产生了深远的影响，也引发了许多新的知识产权保护问题。如何合法合规地开发、运用人工智能技术，如何做好人工智能企业的知识产权保护和管理，从基础创新保护到转化运用进行全方面、全流程的知识产权体系建设，这些问题均需要在当前法律法规的引导下，结合实际案例和企业经营状况，不断调整和完善，才能逐渐达成目标。

（陈茜对本文亦有贡献）

王红燕
合伙人
知识产权部
杭州办公室
+86 571 5662 3968
gracewang@zhonglun.com

涉美欧人工智能业务的知识产权合规要求

变化趋势及应对建议

ARTICLE BY 张鹏 牟雨菲

ChatGPT的横空出世引发了全球的赞誉，人工智能业务已经成为诸多创新型企业的重要赛道。美欧等逐步对人工智能业务提出更高的合规要求，其中就包括知识产权合规的重要内容。特别是，欧盟《人工智能法案》立法程序的加快推进，美国参议院司法委员会知识产权分委员会"人工智能和知识产权"组织专门听证会等。本文结合上述情况对涉美欧人工智能业务的知识产权合规要求变化趋势进行分析，并给出企业应对上述趋势的相关建议。

“人类在进入到21世纪的三个关键时间点，相继出现了三个互相联系又略有区别的新时代，即网络社会时代、大数据时代、人工智能时代，三者共同构成了新的社会时代。”[1]人工智能技术发展影响着人类社会的方方面面，将改变甚至颠覆人类现存的生产工作和交往方式，由此出现一个以新的技术结构支撑新的社会结构的人类新时代[2]。

知识产权制度是受到人工智能技术发展影响最为直接的法律制度之一，各国均对此高度关注[3]。2023年8月15日，我国国家互联网信息办公室、国家发展和改革委员会等公布施行《生成式人工智能服务管理暂行办法》，该办法为AIGC服务的知识产权保护提供了明确指导[4]。2023年7月12日、5月17日，美国国会连续组织两次听证会，分别以“人工智能和知识产权：第一部分 人工智能与版权法的互操作性”、“人工智能和知识产权：第二部分 版权制度的完善”为主题进行了讨论。[5]2023年6月14日，欧盟议会以 499 票赞成、28 票反对和 93 票弃权通过了《人工智能法案》的谈判授权草案，意味该法案即将进入最终谈判阶段，亦涉及知识产权相关规则的完善。[6]本文结合欧盟《人工智能法案》知识产权部分，和美国参议院司法委员会知识产权分委员会召开关于人工智能和知识产权关系的听证会的相关进展，探讨人工智能技术发展对知识产权制度可能的挑战，以及我国法律实践未来的可能回应。

1.何哲：“通向人工智能时代——兼论美国人工智能战略方向及对中国人工智能战略的借鉴”【J】，载于《电子政务》2016年第12期，第2-10页。
2.吴汉东：“人工智能时代的制度安排与法律规制”【J】，载于《法律科学》2017年第5期，第128-136页。
3.参见王红燕著：《中国人工智能合规建设与知识产权法律实务》【M】，北京：中国法制出版社2023年1月版，第81页。
4.参见中伦视界公众号“跨越AIGC合规上市之路”系列文章。
5.*Artificial Intelligence and Intellectual Property: Part I — Interoperability of AI and Copyright Law*，https://judiciary.house.gov/committee-activity/hearings/artificial-intelligence-and-intellectual-property-part-i
Artificial Intelligence and Intellectual Property – Part II: Copyright,
https://www.judiciary.senate.gov/artificial-intelligence-and-intellectual-property_part-ii-copyright
6.*MEPs ready to negotiate first-ever rules for safe and transparent AI,*
https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai

/PART 001

美国人工智能和知识产权听证会的基本情况概述

美国国会于2023年7月12日、5月17日连续组织两次听证会，分别就"人工智能和知识产权：第一部分　人工智能与版权法的互操作性"、"人工智能和知识产权：第二部分 版权制度的完善"进行了讨论：

第一部分听证会于5月17日举办，讨论了版权保护作品在生成式人工智能模型训练中的使用，使用生成式人工智能辅助创作的作品的版权保护，以及生成式人工智能对创作者和创意产业的经济影响等话题。于2023年7月12日进行的第二部分听证会则探讨关于人工智能生成内容涉及的一系列版权问题，包括生成式人工智能对于传统版权法体系的挑战以及相应的应对措施等。

上述听证会具有较强的行业影响力，其中就人工智能知识产权问题的讨论具有较高的实务价值以及权威性。首先，两次听证会的参加人员包括律师、计算机专家、法学家、软件公司从业人员、知名歌手等，具有较高的代表性。其次，两次听证会中多数意见认为不应当对生成式人工智能输出内容作为版权保护客体。例如，PLUS Coalition总裁兼CEO，专业摄影师Jeffrey Sedlik提出，版权法是为人类的原创表达提供保护的，而生成式人工智能输出的图像并非人类原创表达，不应获得版权保护的资格。[7]格莱美提名音乐艺术家Dan Navarro在书面证词中提到其协助发起了一项旨在倡导人工智能时代创作者权利的"人类艺术性运动"。[8]多名艺术家在听证会发言中均强调了"人类艺术性运动"提出的"七项高层次指导原则"[9]。该原则是音乐创作者针对科技发展所自发组织而形成的，关于科技与创作共存的基本原则，强调：①音乐家们将利用人工智能这一最新技术来做伟大的创新。②人类创造的作品在我们的生活中仍将是

7.https://docs.house.gov/meetings/JU/JU03/20230517/115951/HHRG-118-JU03-Wstate-SedlikJ-20230517.pdf
8.https://docs.house.gov/meetings/JU/JU03/20230517/115951/HHRG-118-JU03-Wstate-NavarroD-20230517.pdf
9.https://www.humanartistrycampaign.com

必不可少的。③出于人工智能目的使用受版权保护的作品，以及使用专业表演者的声音和肖像，需要得到许可并提供合理经济报偿。④政府不应创设新的版权或其他知识产权豁免。⑤版权法应该只保护人类智力创造的独特价值。⑥可信度和透明度对人工智能的成功和保护创作者至关重要。⑦创造者必须有一席之地，而不仅仅是只有开发者。

两次听证会均提及著作权合理使用制度在人工智能发展中的重要作用：

在生成式人工智能训练模型中，版权法的合理使用原则对使用版权内容产生何种影响是这两次听证会的主要焦点。前美国版权局总法律顾问Sy Damle指出，版权合理使用制度是平衡人工智能领域竞争利益的最佳方式。[10]Jeffrey Sedlik认为，针对人工智能摄取受版权保护的创作成果，合理使用仅能作为人工智能开发者的一项肯定性辩护，而非一项权利。对合理使用的判断是针对具体事实的调查，必须在个案的基础上进行。埃默里大学法学院法学、人工智能、机器学习和数据科学教授Matthew Sag表示，根据合理使用原则，AI生成平台本身将被认为是一种非表达性使用，这一观点可类推至关于搜索引擎或抄袭检测程序运行属于合理使用的相关司法认定。[11]

两次听证会提及创设联邦层面的公开权或反模仿权被认为是对生成式人工智能模仿艺术风格问题的解决路径：

UMG（环球音乐集团）Jeffrey Harleston认为，可以通过一项联邦权利来解决对唱片公司旗下音乐艺术家进行大肆模仿的深度伪造问题，相较于目前州际公开权下的法律框架，联邦层面的权利可以提供更多的一致性以及更宽的覆盖范围。目前，各州的公开权法规为根据类似的联邦权利所产生的诉因（例如，上面提到的对音乐艺术家的深度伪造）提供了一些潜在的救济。[12]

10.https://docs.house.gov/meetings/JU/JU03/20230517/115951/HHRG-118-JU03-Wstate-DamleS-20230517.pdf
11.https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_sag.pdf
12.https://www.judiciary.senate.gov/imo/media/doc/2023-07-12_pm_-_testimony_-_harleston1.pdf

/PART 002

欧盟《人工智能法案》谈判授权草案基本情况概述

《人工智能法案》于2021年4月由欧盟委员会提出草案，旨在确保在欧洲开发和使用的人工智能完全符合欧盟的权利和价值观，包括人类监督、安全、隐私、透明度、非歧视以及社会和环境福祉。欧洲议会议员对文本进行了几项重大修改，包括人工智能定义、禁止行为的清单，并且对高风险人工智能的场景做了进一步明确，维持了通用人工智能相关主体的义务。该法案采用基于风险的方法，将人工智能应用归类为："无法被接受的风险应用""高风险应用""有限风险应用"以及"无风险或轻微风险"应用。法案严格禁止"对人类安全造成不可接受风险的人工智能系统"，包括有目的地操纵技术、利用人性弱点或根据行为、社会地位和个人特征等进行评价的系统等。受管制的某些人工智能应用被视为高风险（例如医疗装置、机械），应当根据有关的行业法规接受第三方评估，合格后才能投入市场。其类别主要包括自然人的生物识别和分类、执法、司法和民主进程等方面的应用。最后，大多数人工智能系统不会是高风险的。对于有限风险的人工智能系统，法案设定了透明度义务。欧盟成员国将设立监督机构，确保这些规则得到遵守。一旦获得批准，欧盟《人工智能法案》将有可能成为全世界首部有关人工智能的法规。[13]

从法案进程角度而言，相较之今年早期进入三方会谈程序的《欧盟数据法案》（EU Data Act），《人工智能法案》位于欧盟法案颁布更为早期的阶段。该法案适用普通立法程序，由欧盟委员会提出，欧洲议会（代表欧盟公众）以及欧盟理事会（代表欧盟各成员国政府）共同商议修订并达成一致后方可准予颁布。[14]《数据法案》已经进行了几次三方会谈，但《人工智能法案》还未开

13.https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
14.https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=celex:52021PC0206

始三方商谈程序，预计仍需相当一段时间，欧盟才可以确认由欧盟委员会、欧盟议会以及欧盟理事会三方达成共识的最终修订版本。各方会分别内部商议确定各自的修订版本，随后由欧盟委员会组织进行多次三方会谈，确认最终版本。

2022年12月，欧盟理事会发布了其修订版本，修订的重点在于：限缩人工智能定义、将社会评级的AI禁止拓展至私人主体、细化"高风险"AI界定（不禁用但受限）、细化高风险AI的义务、增加针对多重用途的AI的相关规定、明确该法案的适用范围、简化义务机制、细化AI的公开透明义务、修改关于激励创新的相关规定等。[15]

2023年6月14日，欧洲议会通过了《人工智能法案》的谈判授权草案。欧洲议会对于该法案的修订主要集中在：将AI的定义与OECD相关规定协调一致，大幅修改增加禁用AI列表，将高风险AI的定义限缩到必须造成"显著风险"，增加高风险AI的基础权审查机制，增加一般用途AI的相关规定。同时，一般基础AI（例如ChatGPT）也将被赋予公开透明义务，所有基础AI模型的创作者需向下级运营商提供一切其遵守该法案所需要的信息，并且建议设立欧盟级别的AI办公室。欧盟颁布《人工智能法案》不仅将促进人工智能的应用发展，并旨在确保涉欧人工智能应用更有助于实现健康安全、基本权利保障等立法目的的实现，同时旨在将鼓励创新这一重要立法目的予以落实。[16]

一方面，基于与美国版权合理使用制度扩张类似的思路，将人工智能系统投放市场或投入使用之前有关该系统的研究、测试和开发活动排除在《人工智能法案》规制的范围之外。《人工智能法案》第2条第5d项提出，该法案不适用于在人工智能系统投放市场之前的研究、测试和开发活动。委员会有权通过授权法案，明确规定这一豁免，以防止其现有和潜在的滥用。人工智能办公室为研究和开发的管理提供指导，其目的也是为了协调国家监督机构的应用。

15.https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence
16.ibid.

另一方面，《人工智能法案》提出保密义务的要求。《人工智能法案》第28条"供应者、分发者、进口者、部署者或其他第三方在人工智能价值链上的责任"的第2b项提出，"就本条而言，商业秘密应得到保护，并且只能在事先根据（欧盟）2016/943号指令采取所有具体的必要措施来维护其机密性，特别是对第三方而言。必要时，可商定适当的技术和组织安排以保护知识产权或商业秘密。"《人工智能法案》第64条规定，"获取数据和文件"要求人工智能开发者及提供者应允许国家监督机构通过其他适当的技术手段和工具，全面查阅提供者或部署者使用的训练、验证和测试数据集。并在满足相关条件下，应允许国家监督机构查阅人工智能系统的训练用和训练后模型，包括其相关模型参数。获得的所有符合法案第70条的信息应被视为机密信息，应遵守现有的欧盟知识产权和商业秘密保护法，并应在调查结束后予以删除。

自2023年6月起，欧盟委员会、欧盟议会以及欧盟理事会已经于6月、9月与10月就《人工智能法案》进行多次会谈。[17]2023年11月10日，欧洲三大经济体，德、法、意通过一份联合文件明确反对法案整体对人工智能应用的层级分类模式（tired model）以及通用人工智能应用（如ChatGPT）的法律意义上的强制义务，而主张对于此类通用人工智能应用，应仅保留无强制效力的实践准则模式（Codes of Conduct）。[18]2023年11月19日，欧盟委员会发布了该法案最新可能版本，其维持了层级分类模式，对通用人工智能应用（如ChatGPT）进行了更详细的应用分层方法，略微放宽了此类应用的合规义务，并进一步明确实践准则。[19]

2023年12月9日，欧盟理事会主席国和欧洲议会谈判代表就该法案达成

17.https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence
18.https://www.euractiv.com/section/artificial-intelligence/news/france-germany-italy-push-for-mandatory-self-regulation-for-foundation-models-in-eus-ai-law/
19.https://www.euractiv.com/section/artificial-intelligence/news/ai-act-eu-commission-attempts-to-revive-tiered-approach-shifting-to-general-purpose-ai/

了临时协议。[20]此次修改的基本要点包括：（1）进一步确立完善通用型人工智能的监管模式，重点强调高风险或可能存在较高系统性风险的通用型人工智能：（2）就执法部门于公共场所就生物识别系统的使用，原法案版本为完全禁止，临时协议中改为受监管条件下的使用；（2）增加禁止社会评级系统以及任何操纵或利用用户个人基本权益的人工智能应用；（3）增强法案可执行性，例如进一步保障个人的异议权和获得解释权；（4）确认违反本法案强制性规定的惩罚性赔偿从3500万欧元(占全球营业额的7%)到750万欧元(占全球营业额的1.5%)不等；（5）对"高风险应用"（即仅次于禁止性应用的最高监管层级）开发者创设人工智能系统投入使用前的基本权利影响评估报告义务。除上述关键要点以外，双方还就法案细节进行了讨论修订，例如定义以及适用范围等部分。[21]总体而言，本次临时协议的修订进一步巩固了层级式监管模式，细化调整了禁止性应用以及高风险应用的范围。最主要的是，随着通用型人工智能的应用发展，近几次修订的要点均集中在提高和确立通用型人工智能的监管必要性以及监管程度等问题。

可以说，此次通过临时协议是为了确保能够在2024年5月欧洲议会选举前顺利通过这项立法。下一步，在临时协议后，双方将继续开展技术层面的协调工作以最终确定法规细节内容，随后由欧盟理事会主席国提交欧盟理事会成员国代表（Coreper）进行批准。在该法案正式通过欧盟立法委员会前，法案文本还需得到欧盟理事会以及欧洲议会的最终确认。一旦通过，还将需要两年时间才能完全生效。

20.https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai

21.https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/

/PART 003

涉美欧人工智能业务的知识产权合规应对建议

如前所述，美欧积极加快人工智能知识产权规则的建设，这对我们开展涉美欧人工智能业务企业的知识产权合规提出新的更高要求。为此，我们建议：

一是，积极追踪人工智能法律法规的最新进展，为业务合规提供基础支撑。如前所述，各国积极探索建构人工智能法律法规，对人工智能产业发展进行规制，人工智能领域立法和政策出台的速度明显加快。在这样的背景下，2017年以来，美国、欧盟、中国、加拿大、日本、新加坡等国家或地区陆续发布人工智能发展和治理的法律法规准则，知识产权是其重要内容。国务院办公厅印发的《国务院2023年度立法工作计划》明确部署，"在实施科教兴国战略、推进文化自信自强方面……预备提请全国人大常委会审议……人工智能法草案。"这就需要我们根据业务发展情况，积极追踪人工智能法律法规的最新进展，对业务合规标准实时更新，为业务合规提供基础支撑。

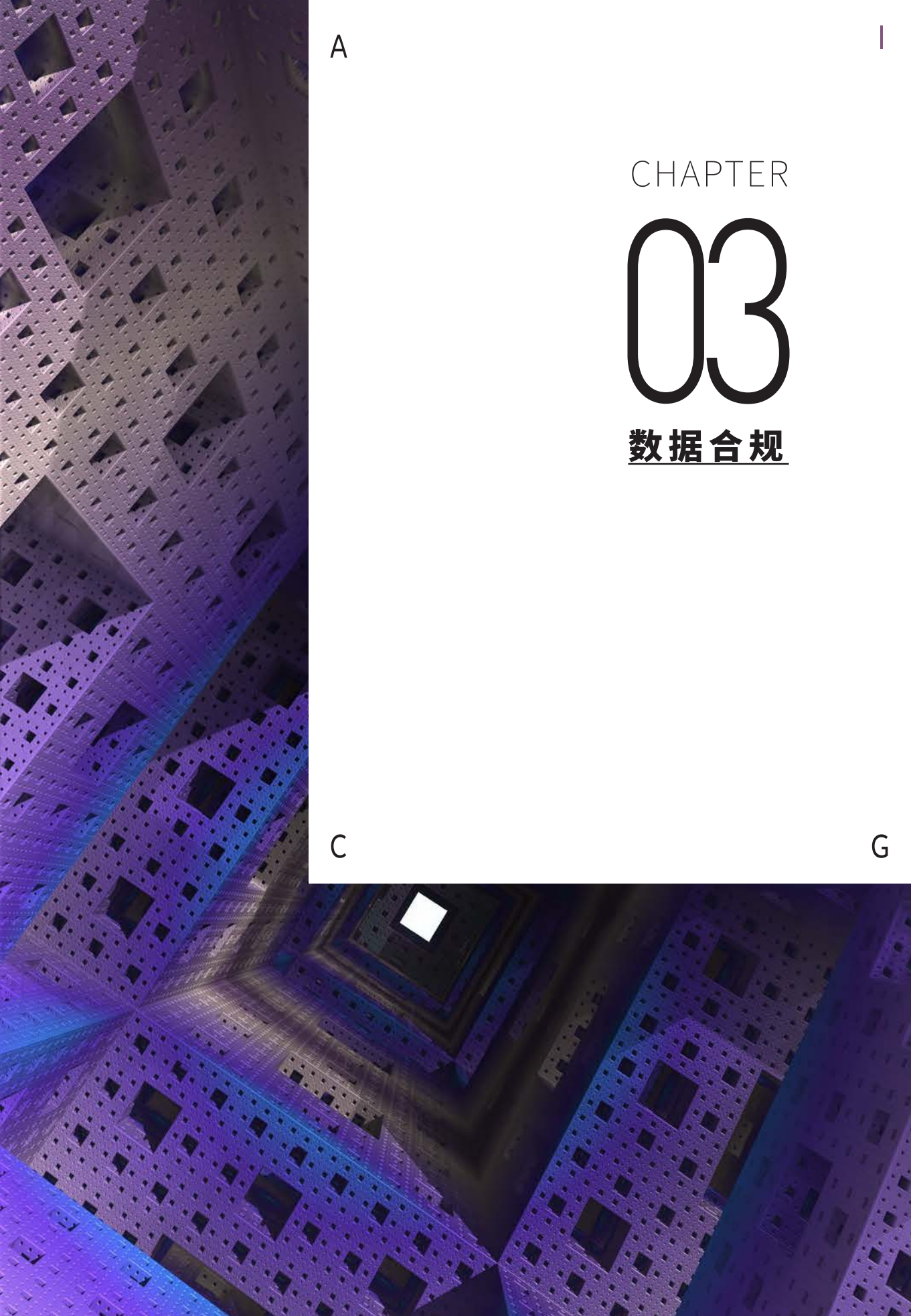
二是，强化人工智能领域的知识产权布局，为业务创新提供权利保护。人工智能作为当前最为尖端的科技成果，对于专利制度的挑战是全方面的，既包括人工智能技术本体的专利法律保护、人工智能发明成果的专利法律规制，还涉及人工智能应用工具的专利法律影响问题。这其中，人工智能技术本体的可专利性问题，也即人工智能技术能否纳入专利法保护客体范围，是上述专利制度面临的首要问题；亦即，随着人工智能技术专利申请数量的大幅增加，其是否构成专利法保护的客体成为法律实践中的重要争议点。只有先明确哪些人工智能技术能纳入专利法保护客体范围，才能进一步明确这些人工智能技术的新颖性和创造性判定、专利文件撰写要求、权利归属、保护范围、侵权判定、侵权救济等。人工智能技术可专利性需要考虑如何对基础算法的保护需求加以回应，以及如何在感知层、认知层、应用层的技术方案中实现对基础算法的实质保护。

对于人工智能技术是否属于专利法保护客体的审查标准，欧盟采用"技术属性测试法"，美国采用"拟制现有技术排除测试法"，我国采用的标准近似于欧盟"技术属性测试法"和美国"拟制现有技术排除测试法"二者的交集，亦即需要满足两个方面的要求才能属于我国专利法的保护客体[22]。进而，AIGC服务提供者应将算法与实际解决的技术问题相结合，体现算法步骤的执行能够采用具有自然规律的技术手段，解决具体技术领域的具体技术问题，并形成具体的技术效果，进而在算法设计中将各个特征与参数与解决该具体技术问题的特征和参数紧密关联，使之成为可专利的新发明[23]。

三是，强化人工智能技术进出口的合规分析，同时防范知识产权侵权风险。从企业知识产权合规的角度而言，一方面，需要加强知识产权布局保护好自身的创新；另一方面，需要尊重他人的知识产权，防范知识产权侵权风险。针对人工智能相关技术的跨境技术交易，需要加强进出口合规的分析，并且在条款设计中对技术进出口合规的风险作出有效管控[24]。同时，人工智能技术的知识产权诉讼具有自身的一些特点，需要结合产业的情况进行分析[25]。对于涉美欧人工智能业务的企业来说，应当充分认识到人工智能技术知识产权诉讼的特点，并且将其运用到风险防控之中。

22.张鹏："人工智能技术的可专利性探析（上）（下）"【EB/OL】，载于中伦视界公众号2021年8月18日和2021年8月20日。
23.蔡鹏、张鹏、胡云浪、于丽君："跨越AIGC合规上市之路（四）：知识产权合规篇"【EB/OL】，载于中伦视界公众号2023年8月15日。
24.参见张鹏、牟雨菲："工程软件源代码跨境知识产权交易实务六问"【EB/OL】，载于中伦视界公众号2023年4月25日；张鹏："跨境知识产权交易的框架设计与关键条款安排"【J】，载于《中国律师》2021年第5期，第46-47页等。
25.参见中伦代理人工智能领域"智能语音机器人"专利案件【EB/OL】，载于https://www.zhonglun.com/Content/2022/08-16/1522211902.html（2023年8月18日最后访问）

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



CHAPTER

03

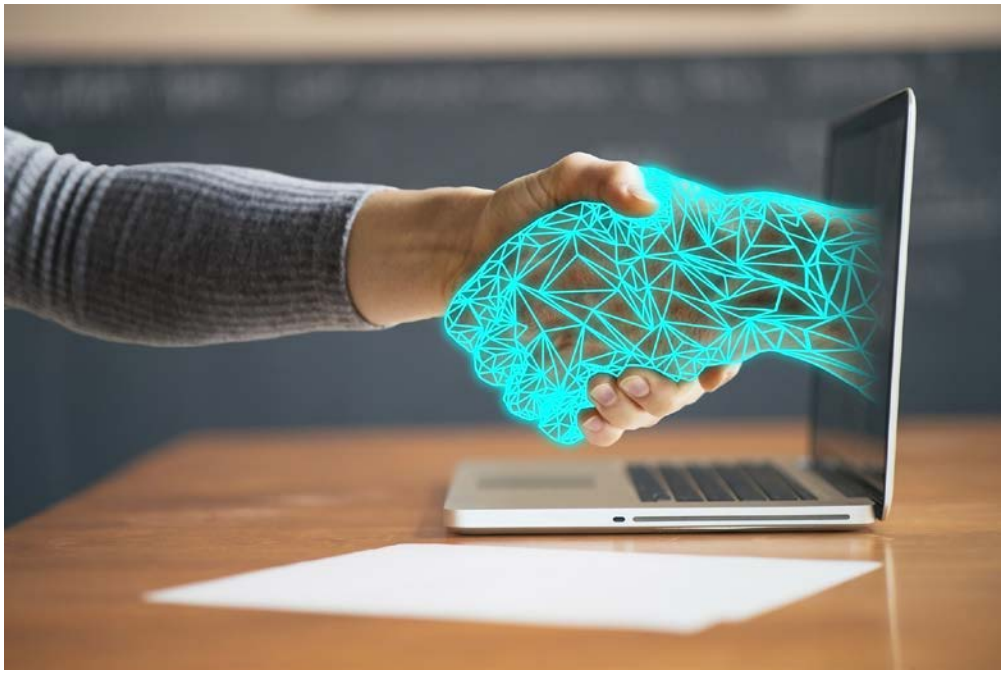
数据合规

全景透视生成式人工智能的法律挑战(二)：

数据合规挑战与路径

ARTICLE BY 陈际红 吴佳蔚 陈煜烺

AIGC全生命周期所牵涉的数据合规问题非常复杂，主要阶段包括模型训练、应用运行和模型优化，通常还涉及AIGC开发者、服务提供者、服务使用者等多方主体。本篇由面及点，从整体的视角提出管理AIGC数据合规风险的方法论，并就其中的数据合规焦点问题进行讨论。



/PART 001

风险识别：三维度的AIGC数据合规风险管理框架

世界各国都在关注和研究人工智能的风险，美国国家标准技术研究院（National Institute of Standards and Technology）于2023年1月发布人工智能风险管理框架（AI Risk Management Framework，AI RMF）[1]，该框架明确了AI系统的生命周期及各阶段的参与者、关键维度，为人工智能的风险管理提供了系统化的评估路径。我们参考该AI RMF，结合AIGC业务特点提出三维度的数据合规风险管理框架。

AIGC的生命周期大体可包括模型训练、应用运行和模型优化三个阶段，各阶段涉及各方主体的数据处理活动及其风险因素如下：

1.See NIST AI 100-1 Artificial Intelligence Risk Management Framework (AI RMF 1.0). https://doi.org/10.6028/NIST.AI.100-1.

阶段	模型训练	应用运行	模型优化
主要活动	立项设计，数据采集 数据清洗，数据标注 模型训练，模型验证	AIGC服务使用者输入内容，AIGC服务提供者生成内容	利用应用运行阶段采集的数据开展模型优化
主体[2]	AIGC开发者[3] 数据主体/数据提供方	AIGC开发者 AIGC服务提供者 AIGC服务使用者	AIGC开发者 AIGC服务提供者 AIGC服务使用者
核心风险要素[4]	隐私性（个人信息、商业秘密）与合法性 数据质量 可靠性与稳健性	隐私性与合法性 透明性与可解释性 准确性与公平性 应用风险 信息内容监管 信息安全	隐私性与合法性

（一）模型训练阶段

主要活动及典型风险：模型的成熟度及生成内容的质量都与训练数据高度相关，故此阶段涉及大量数据收集，并对此等数据进行清洗和标注（tokenization）后用于模型训练和验证。数据清洗、标注及模型训练存在内部性，需

2.除本表所列主体外，实践中AIGC开发者/服务提供者还可能引入其他系统/产品供应商，以支撑其模型训练或业务运营的软件/硬件环境、技术支持等。

3.《生成式人工智能服务管理暂行办法》第二十二条第（二）项所定义的"AIGC服务提供者"概念实际上已涵盖了"通过提供可编程接口等方式提供AIGC服务"的主体，为便于区分主体以厘清各方责任，本文针对此等"提供可编程接口的技术开发者"与"服务提供者"并非同一主体时，使用"AIGC开发者"的概念。

4.事实上，AIGC技术的模型训练、应用运行和模型优化等各个阶段涉及的数据合规风险要素包括：隐私性（个人信息、商业秘密）与合法性、可靠性与稳健性、透明性与可解释性、准确性与公平性、应用风险、信息内容监管、信息安全（完整性、机密性、可用性），本表仅列举各阶段所涉及的核心风险要素。

要重点关注AIGC模型（及所用于训练的数据）的可靠性与稳健性以及数据质量（真实性、准确性、客观性、多样性[5]），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；而数据收集的风险则需关注数据源合法性，《暂行办法》第七条即要求AIGC服务提供者"使用具有合法来源的数据和基础模型"，典型数据收集形式及合规风险包括：

✓ 采取网络爬虫等形式爬取数据，具体分析见下；

✓ 收集已合法公开的公共数据，例如收集根据各地政务数据共享开放条例所合法公开的数据；

✓ 直接收集：直接面向人信息主体收集数据，如收集其个人信息，则需满足个人信息合规的要求；

✓ 间接收集：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；

✓ 合成数据（计算机模拟生成的数据），应主要关注数据质量。

主体：AIGC开发者及数据主体/数据提供方。

（二）应用运行阶段

主要活动及典型风险：将AIGC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。具体而言：

✓ AIGC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；

✓ 而AIGC生成内容时也可能存在隐私风险及数据泄露等风险；

✓ 就AIGC服务提供者处理、分析数据而言，可能存在超目的处理、未就数

5.《生成式人工智能服务管理暂行办法》第七条第（四）项。

据共享、数据跨境进行充分告知并取得有效同意等风险；

✓ 而就对外提供AIGC服务本身，也将面临可靠性与稳健性、透明性与可解释性、准确性与公平性等风险。

主体：AIGC服务提供者及服务使用者。其中既包括AIGC开发者直接面向服务使用者提供服务，也包括通过引入第三方AIGC开发者技术能力进而面向服务使用者提供服务。

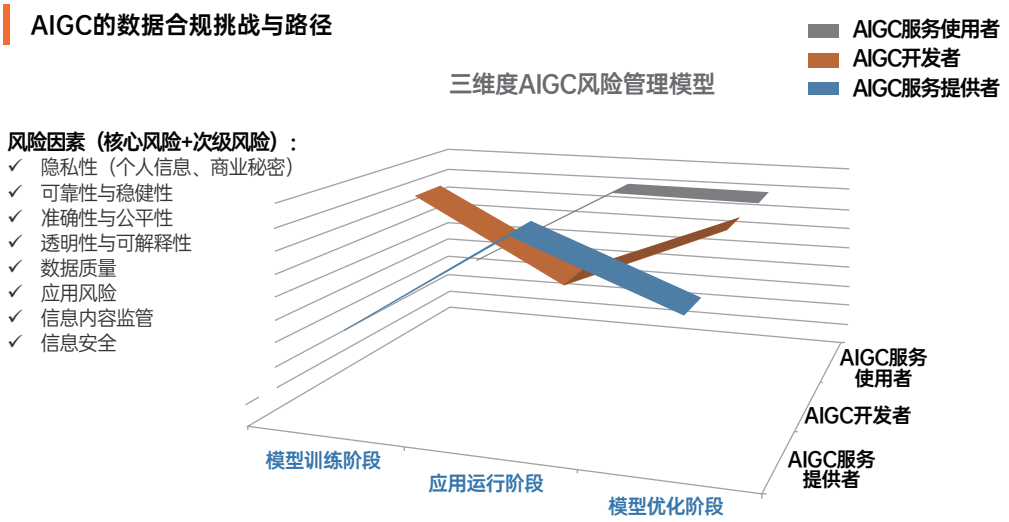
（三）模型优化阶段

主要活动及典型风险：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、准确性与公平性等风险；另一方面，此阶段的外部风险集中在向AIGC服务使用者提供服务时，需明确就此等模型迭代训练等处理活动事先告知AIGC服务使用者并取得有效同意。

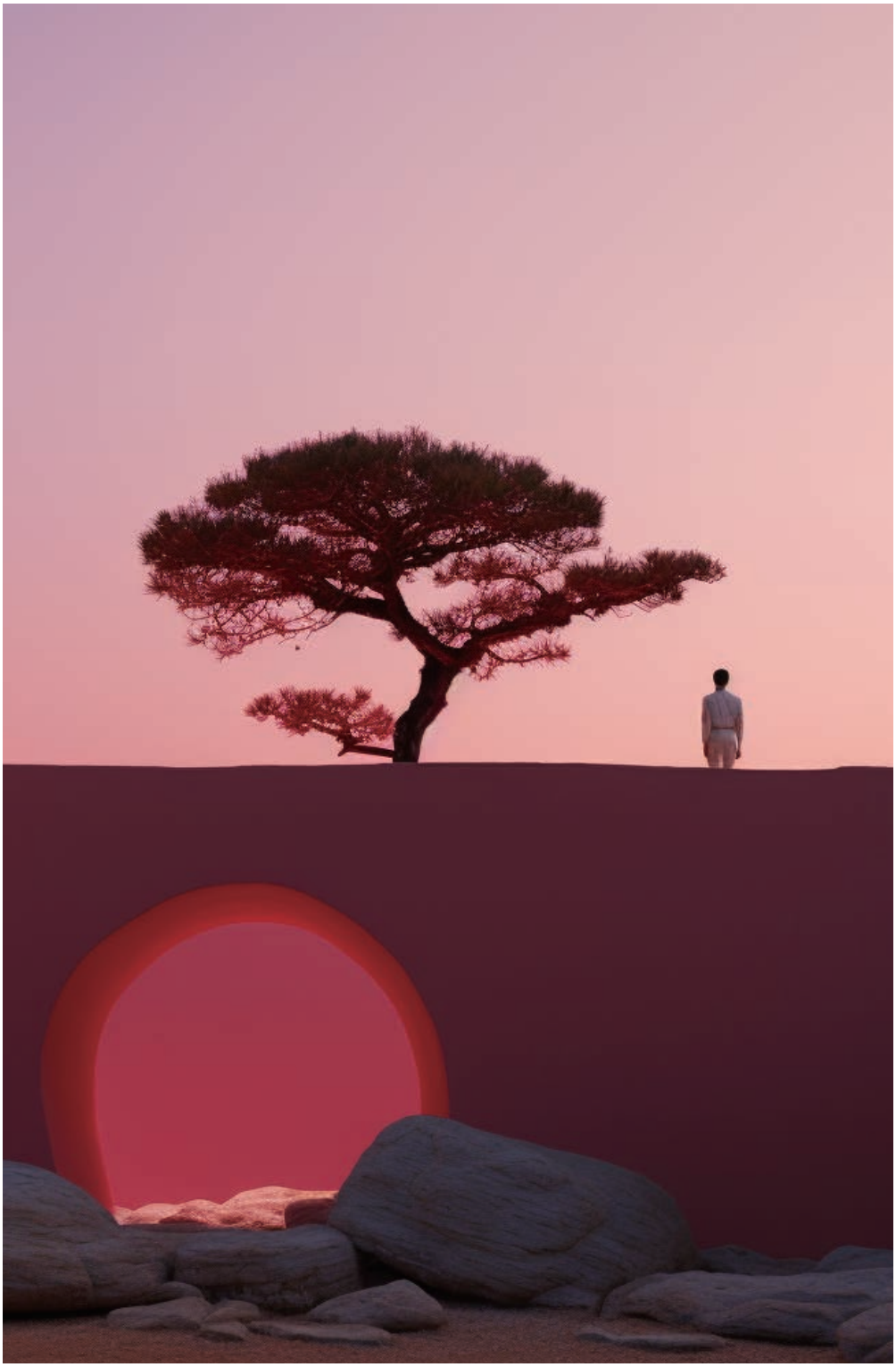
主体：AIGC服务使用者及AIGC开发者。在私有化部署及"API-定制化服务"等模式下，集成AIGC技术的服务提供者亦可能构成责任主体。

三维度的AIGC合规方法论

AIGC的数据合规挑战与路径



注：AIGC服务提供者既可能是集成方，也可能是AIGC开发者直接向用户提供服务。



/PART 002

合规挑战：我国监管框架下AIGC涉及的典型数据合规问题

问题一：如何管理爬取数据的风险

网络爬虫通常是AIGC开发者收集数据并用于模型训练的常见手段。利用网络爬虫抓取数据的法律风险包括：

业务环节	法律风险	法律依据
抓取手段	– **刑事风险**：为越过被爬取网站的反爬虫技术措施而采取反反爬虫技术且获取数据，可能构成"非法获取计算机信息系统数据罪"[6]；如果因抓取行为造成被爬网站服务器不能正常运行，可能构成"破坏计算机信息系统罪"； – **不正当竞争**：如果违反目标网站Robots协议，或通过IP代理、伪造UA（User-Agent）等形式，或其他反反爬虫措施实施抓取，扰乱被爬网站的正常运营，则可能构成《反不正当竞争法》下的"妨碍、破坏其他经营者、的网络产品或者服务正常运行"[7]的不正当竞争行为； – **著作权保护**：如果未经权利人许可，故意避开或者破坏权利人设定的技术措施[8]，以抓取他人享有著作权的作品，可能承担民事侵权责任。	《刑法》第285、286条 《反不正当竞争法》第12.2条 《著作权法》第49.2、53（6）条

6.司法实践中，诸如IP代理等行为均可能被认定为侵入性行为，典型案例包括：（2017）京0108刑初2384号、（2017）浙0110刑初664号、（2019）闽08刑终223号、（2016）浙0681刑初1102号、（2016）沪0115刑初2220号、（2016）浙0602刑初1145号。

7.《反不正当竞争法》第十二条第二款规定，利用技术手段，通过影响用户选择或者其他方式，实施妨碍、破坏其他经营者合法提供的网络产品或者服务正常运行的行为，构成不正当竞争。

8.具体而言，技术措施包括两种：1）访问控制技术措施（又称防止未经许可获得作品的技术措施），该类技术措施是通过设置口令等手段限制他人阅读、欣赏文学艺术作品或者运行计算机软件；2）保护版权人专有权利的技术措施（又称保护版权专有权利的技术措施），即防止对作品进行非法复制、发行等的技术措施。

业务环节	法律风险	法律依据
抓取内容	**– 个人信息保护：**如果抓取内容构成个人信息，需满足透明性要求并具备合法性基础，避免爬取未合法公开的个人信息，否则可能构成"违法收集个人信息"； **– 著作权保护：**如果未经权利人许可，抓取他人享有著作权的作品，则此等抓取作品至自身数据库的行为可能构成侵犯复制权[9]。	《个人信息保护法》第13、17条 《著作权法》第10条
抓取数据的后续使用	**– 个人信息保护：**如果抓取已公开的个人信息并用于后续模型训练，则需确保满足法律关于处理已公开个人信息的要求，例如需在合理范围内处理且在对其权益有重大影响时取得同意[10]； **– 著作权保护：**如果未经权利人许可，抓取他人享有著作权的作品，进行分析、处理、演绎，并在后续输出成果中展示，亦可能侵犯他人著作权； **– 不正当竞争：**如果抓取数据后的模型训练、提供AIGC服务行为被视为"对被爬网站构成实质性替代"[11]，或者"爬取'衍生数据'构成'搭便车'"[12]，则可能受到《反不正当竞争法》的规制。	《个人信息保护法》第27条 《著作权法》第10条 《反不正当竞争法》第2、12条

9.关于合理使用的讨论，请见"（二）知识产权篇"。

10.《个人信息保护法》第二十七条　个人信息处理者可以在合理的范围内处理个人自行公开或者其他已经合法公开的个人信息；个人明确拒绝的除外。个人信息处理者处理已公开的个人信息，对个人权益有重大影响的，应当依照本法规定取得个人同意。

11.例如点评诉百度案、点评诉爱帮案，案件具体情况参见（2016）沪73民终242号民事判决书、（2010）海民初字第24463号民事判决书。

12.例如"淘宝诉美景案"，案件具体情况参见（2018）浙01民终7312号民事判决书。

值得注意的是，尽管目前司法实践中大量案件通过《反不正当竞争法》规制网络运营者之间的爬取行为及使用行为，**但《反不正当竞争法》的适用前提是网络运营者之间存在竞争关系，**这就涉及对AIGC所涉及服务性质的认定，并结合二者是否存在相关市场、爬取行为是否可能对被爬网络运营者现实或潜在经济利益造成损害、是否争夺其交易机会或竞争利益（例如用户流量）等因素，具体仍有待于观察司法实践如何认定。

问题二：如何管理数据源风险

《暂行办法》第七条要求AIGC服务提供者（含开发者）"使用合法来源的数据及基础模型"，不仅增加了对于数据的合法性要求，还强调了模型的合法性要求[13]。实践中，AIGC开发者可能通过引入第三方（数据提供方）数据以训练自身模型，也可能通过联合建模（但合作方数据不会提供至AIGC开发者）等形式开展。就AIGC开发者向数据提供方间接收集数据以用于模型训练的场景，核心在于确保数据源的合法合规。企业可对此等数据提供方采取的合规管理措施包括：

– 审查数据提供方是否具备提供数据的法律依据（legal basis），如涉及模型合作的，需审查模型是否具备合法来源。针对个人信息，应审查其是否具备提供个人信息的合法性基础[14]；针对非个人信息，应确保所提供的数据不存在权属瑕疵；

– 与数据提供方签署数据处理协议或在业务合同中加入数据保护条款，明确各方权利义务及责任，以协议形式要求该数据提供方承诺数据源的合法合规；

– 针对数据提供方开展合规尽调，就其数据保护层面的安全能力、组织管

13.部分域外案例中，Clearview AI 和OpenAI的数据来源以及模型的合法性均受到法律挑战。

14.《个人信息保护法》第十三条规定了处理个人信息的合法性基础，包括取得个人同意、基于履行合同或人力资源管理所必需、基于履行法定义务所必需等。

理、操作规程等进行调查，避免因数据提供方的整体运营及业务模式不合规而影响合作持续开展；

– 要求数据提供方配合定期开展合规审计等。

问题三：AIGC应用运行及模型优化阶段典型个人信息合规问题

OpenAI在其官网公布的ChatGPT隐私政策中披露，其产品运营过程中会收集用户账户信息、对话内容（如含个人信息）、交流信息、社交媒体信息，以及日志信息、使用情况、设备信息、Cookies等个人信息，并表明其可能会利用AIGC服务使用者所提供的数据以开展模型优化和改进[15]。《暂行办法》第十一条提出AIGC服务提供者的个人信息保护要求，强调"必要性"，以及不得"非法留存"和"非法提供"，在我国法律框架下，AIGC开发者/服务提供者所面临的典型个人信息合规问题包括：

– **透明性及合法性要求。**《个人信息保护法》要求处理个人信息应告知AIGC服务使用者个人信息处理规则（例如通过隐私政策）并具备合法性基础（例如以勾选框等形式取得同意，或系为AIGC服务使用者提供服务所必需）。实践中，通常应当由直接面向AIGC服务使用者提供服务的AIGC服务提供者承担此等义务。

– **个人信息跨境。**无论是境外AIGC开发者直接面向境内AIGC服务使用者提供服务，还是AIGC服务提供者接入境外AIGC技术API接口后向境内AIGC服务使用者提供服务，均可能涉及将AIGC服务使用者个人信息传输至境外。目前《个人信息保护法》《数据出境安全评估办法》《个人信息出境标准办法》均对个人信息出境提出合规要求，相关方应准确识别个人信息出境场景，选择出境合法机制（标准合同/安全评估/认证），开展个人信息保护影响评估（PIA），于隐私政策等文本中披露跨境情形，并具备相应合法性基础。《暂

15.https://openai.com/policies/privacy-policy，最后访问日期：2023年9月12日。

行办法》第四章"监督检查和法律责任"第二十条特别提及来源于境外AIGC服务的场景，可以预见此等涉及跨境的AIGC服务将面临监管的重点关注。

– **将AIGC服务使用者输入数据用于模型优化。**AIGC服务使用者可期待的个人信息处理目的可能并不包括将其数据用于非本次服务以外的目的，例如模型优化，因此将面临更高的用户权益保护风险。目前Open AI的做法是在隐私政策中披露该情况，并就Non-API模式采取默认同意用于模型优化、除非要求退出（opt - out）的机制，就API模式则采取默认拒绝用于模型优化、除非主动提供（opt - in）的机制[16]。国内大部分大模型则强调，会对输入内容中个人信息进行去标识化或匿名化基础上，将输入内容用于模型训练、优化[17]。在《个人信息保护法》的框架下，如果无法做到完全的匿名化处理，应对模型优化目的处理取得同意，且应采取opt - in的同意机制。

问题四：如何处理AIGC数据泄露风险？

据媒体此前报道，韩国某头部科技企业内部发生三起涉及ChatGPT误用与滥用案例，包括两起"设备信息泄露"和一起"会议内容泄露"。报道称，半导体设备测量资料、产品良率等内容或已被存入ChatGPT学习资料库中，随时

16.详情见Open AI: How your data is used to improve model performance, https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance，最后访问日期：2023年9月12日。

17.**文心一言**："我们将根据相关法律法规的要求通过技术手段对个人信息进行必要的去标识化或匿名化处理，处理后的信息将无法精确识别到特定个人信息主体。请您了解并同意，在不透露您个人信息且不违背相关法律法规的前提下，我们有权对用户数据进行分析并予以利用，包括但不限于使用技术处理后的对话信息提高文心一言对您输入内容的理解能力，以便不断改进文心一言的识别和响应的速度和质量，提高文心一言的智能性。"

讯飞星火："根据适用的法律法规，我们可能会对您的个人信息进行技术处理，使得根据该信息无法精确识别到用户个人，并对技术处理后的信息进行匿名化或去标识化的学术研究或统计分析（包括使用匿名化或去标识化后的语音信息进行模型算法的训练；使用您在使用会写功能时输入的对话信息），以便不断改进算法模型的识别和响应的速度和质量、提高会服务的智能性和对您输入内容的理解能力，但我们不会根据您输入的对话内容对您个人进行特定身份的识别。"

百川智能："根据适用的法律法规，我们可能会对您的个人信息进行技术处理，使得根据该信息无法精确识别到用户个人，并对技术处理后的信息进行匿名化的学术研究或统计分析（包括使用您在使用会写功能时输入的对话信息），以便不断改进算法模型的识别和响应的速度和质量、提高会服务的智能性和对您输入内容的理解能力，但我们不会根据您输入的对话内容对您个人进行特定身份的识别。"

面临泄露的风险。而出现这些事故的根源，均是因为员工将涉密内容输入到了ChatGPT。[18]

AIGC所涉及的数据泄露主要包括两类：

– 一类是基于AIGC自身产品/系统发生数据安全事件而导致的数据泄露，这就要求AIGC开发者及服务提供者根据相关规定制定数据安全应急预案，在发生数据安全事件时及时履行上报及通知义务，并对前述第三方系统/产品供应商采取安全层面的合规管理措施。

– 另一类是由于AIGC服务使用者输入数据中含有公司商业秘密或其他敏感数据（个人信息、重要数据等），导致该等数据进入AIGC学习资料库中，进而面临向他人输出内容时泄露此等数据的风险。企业如在内部引入AIGC技术能力时，应对员工的使用规则作出严格限制；而作为向AIGC服务使用者提供服务的AIGC服务提供者，一方面可通过用户协议或其他交互文本提示AIGC服务使用者应避免输入重要数据、他人个人信息等敏感数据，另一方面应采取一定技术手段防止此等因输出内容而导致的数据泄露发生。

问题五：AIGC开发者与服务提供者如何认定数据权属及相关责任

针对由AIGC服务提供者引入第三方AIGC技术（以下称"**集成方**"）以向AIGC服务使用者提供服务的场景，目前合作模式主要有两种，一种是MaaS（Model as a Service，"模型即服务"）模式，集成方通过接入AIGC开发者API从而使用其技术能力，具体包括"API-标准化服务"和可进行模型微调和数据训练的"API-定制化服务"；另一种是私有化部署，即AIGC开发者将其技术能力私有化部署至集成方[19]。由于此种情形下涉及两方主体，进而可能在数据权属、个人信息保护责任及AIGC监管责任等问题上存在潜在争议。

18.https://www.thepaper.cn/newsDetail_forward_22643886，最后访问日期：2023年9月12日。

19.例如，近期开源ChatGLM-6B可以在消费级的显卡上进行本地部署，https://github.com/THUDM/ChatGLM-6B，最后访问日期：2023年9月12日。

– **数据权属。**私有化部署模式下，AIGC开发者通常无法接触、处理数据，故该模式下数据权属取决于AIGC服务提供者与AIGC服务使用者之间的约定（例如用户协议）；而在MaaS模式下，由于AIGC开发者通常有能力接触和处理由AIGC服务使用者输入至集成方界面的数据，故除需考虑集成方与AIGC服务使用者的约定外，还需考量AIGC开发者与服务提供者之间的协议约定，因此，相关方应提前在协议中对此作出安排，例如AIGC开发者是否有权将AIGC服务使用者输入数据纳入训练数据库用于自身模型优化，避免引发权属争议。

– **个人信息保护。**模型训练阶段的个人信息合规问题集中在AIGC开发者面向个人直接收集数据，此种情形其他合法性基础的论证空间有限，告知同意可能是必选项。而对于应用运行乃至模型优化阶段，争议场景仍聚焦在MaaS模式，此模式下存在多种数据流交互的场景，不同处理目的及各类数据流下各方责任将有所差别。目前《个人信息保护法》根据三种数据处理关系进行了责任分配[20]，因此，服务提供者与AIGC开发者应当对于合作模式（核心为处理目的及数据流）进行清晰界定，例如，如果双方约定基于同一目的共同处理AIGC服务使用者输入的数据，且面向AIGC服务使用者提供服务时同时披露了两方主体，则构成共同处理关系，进而需依法承担连带责任。

– **AIGC监管责任认定。**模型训练阶段由于不涉及服务提供者参与，故该阶段监管责任将直接面向AIGC开发者。在考虑开发者与服务提供者的责任分配时，由于开发者控制训练数据和算法，具有履责的技术优势，而服务提供者会面对用户端，系《暂行办法》所规定的直接监管抓手，但由于其不掌握AIGC底层技术，仅有能力对自身可接触的环节负责（例如对基于API接口的模型开展进一步优化训练），《暂行办法》正式稿第七条亦将此前征求意见稿中"应当对生成式人工智能产品的预训练数据、优化训练数据来源的合法性负责"调

20.例如，在《个人信息保护法》项下，委托处理关系通常由委托方对外承担责任，共同处理关系则由双方承担连带责任，对外提供关系则各自承担责任。

整为“应当依法开展预训练、优化训练等训练数据处理活动”，体现出鼓励技术创新的监管态度。据此，无论是私有化部署还是MaaS模式，均建议开发者和集成方通过签订协议等形式，将此等监管义务在法律允许范围内进行明确分配，对各自合规情况进行审计和核查并留存证明材料，以尽到审慎义务。

除上述争议场景外，诸如信息内容责任、数据安全责任、知识产权侵权责任等责任界定问题均可采取类似的认定逻辑，结合具体合作模式、数据流、用户披露情况以及合同约定等综合予以判定。

/PART 003

合规建议：关注不同主体身份的合规责任

AIGC的数据合规是一个大命题，建议相关方以“点面结合”的方式，一方面按照“三维度的数据合规风险管理框架”，并结合《AIGC合规义务清单》对AIGC全生命周期的数据合规风险进行整体识别，同时重点关注业务开展中的典型数据合规问题。基于此，我们为AIGC相关方提供合规建议如下：

首先，针对AIGC开发者：

1)规范数据收集活动，警惕爬虫风险。企业开展爬虫活动时应：(1)爬取公开的、非保密的前台数据，不应爬取非公开的后台数据；(2)避免采用破解密码、伪造设备IP绕过服务器身份校验、IP代理、伪造UA等技术手段以绕过或破解网站采取的保护数据的技术措施；(3)控制爬取的频率、流量等，不得因爬取数据导致被爬网站无法正常响应，甚至瘫痪。

2)提高数据标注及清洗、模型训练阶段的透明性、可解释性及公平性，积极应对监管。模型训练阶段作为AIGC技术开发的开始，企业应采取措施，确保全程可控可见以及可审计和可追溯（透明性），使AIGC所使用的数据、算法、参数和逻辑对输出结果的影响能够被AIGC服务使用者理解（可解释

性），保证所利用数据及开展决策时的公正、中立、不引入偏见和歧视因素（公平性）[21]，由此，也可为应对后续应用运行及模型优化阶段的透明性、可解释性及公平性风险提供支撑。

其次，无论是AIGC开发者还是服务提供者，均应关注：

1)加强第三方管理，提升可靠性与稳健性。数据提供方、系统/产品供应商等第三方作为外部实体，相较于内部管理具有不可控性和责任对立性。建议提前对其系统/产品可靠性与稳健性、合规情况、安全能力展开调查，并通过签订数据处理协议等形式就双方权利义务责任等予以明确，以缓释此等第三方主体的外部合规风险。

2)开展个人信息保护影响评估（PIA），优先整改高风险事项。企业数据合规工作并非一蹴而就，如何在控制成本投入的同时与监管尺度、行业水位拉齐，需要企业做到有的放矢。从近期立法及执法动作来看，透明性、单独同意、个人信息出境、数据处理目的限定（例如是否可用于模型优化）以及个人信息主体行权已然是监管重点。AIGC开发者/服务提供者可对自身个人信息处理活动开展产品维度或业务维度的PIA，并就监管重点关注的高风险事项开展合规自查和整改，之后再逐步推进自身个人信息保护制度建设、操作规程等，建立数据合规的长效机制。

第三，针对AIGC服务提供者：

如涉及集成AIGC开发者技术，应梳理合作模式，厘清数据权属，明晰各方责任。相较于私有化部署模式，MaaS模式下AIGC开发者与集成方的关系将尤为紧密。建议双方应对业务流、目的流、数据流进行清晰界定，并通过签订合作协议，于外部用户协议明确地披露服务提供主体等形式，事先对各方权

21.具体见全国信息安全标准化技术委员会 大数据安全标准特别工作组：《人工智能安全标准化白皮书（2023版）》。

利义务及责任作出安排，避免潜在争议。

防止数据泄露。为防止因AIGC技术不可预测风险而引发的AIGC服务使用者输入数据（尤其是含第三方敏感个人信息、重要数据等）对外泄露，AIGC服务提供者应通过用户协议等对AIGC服务使用者的使用规则作出限制，并事先制定数据泄露事件应急预案，以备不患。

最后，针对AIGC服务使用者：

1)识别高风险应用场景，开展必要性审查并采取针对性措施。鉴于AIGC的应用存在隐私风险及数据泄露风险，建议对使用AIGC应用场景进行必要性审查，以及是否可以采用其他替代方案。此外，针对高敏感场景但有必要采用AIGC的场景，可设计针对性方案，例如采用私有化部署模式，以避免触发数据外泄、跨境传输等风险。

2)规范AIGC使用行为，防止数据泄露。企业如欲在内部引入AIGC工具，建议在员工手册等内部制度中要求其不得输入保密数据等未向外部公开的数据。

3)确保在依据AIGC输出结果采取行动之前进行人工审查。AIGC的技术原理决定其具有准确性和偏见层面的风险，企业应避免依赖AIGC输出结果直接采取行动，而应对其输出结果进行人工审查。

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

浅析人工智能系统训练数据的合规问题

ARTICLE BY 顾萍 詹凯维

ChatGPT让2023年成了人工智能系统（“AI”）的大年。和ChatGPT一样让人感到震惊的是，AI能找到并理解人类认知范围之外的规律，并据此作出解决方案。

如，AI发现了具有显著效果的抗生素——halicin，但是halicin并未具备人类科学家归纳出来的化学特征。科研人员准备了一个包含了两千多个分子的训练数据集，并按照抗菌效果的优劣提前作好标记，让AI在训练过程中可以有的放矢。这是一种经典的机器学习算法——监督学习（supervised learning），即“仅用标注数据进行训练的机器学习”[1]。但是，如果要学习的数据量超出标记能力，AI就需要在大量数据中自动发现其中的规律和联系，这种方式被称为“无监督学习（unsupervised learning）”。比如电商平台的推荐算法：它不关心你具体会买什么商品，它只是在你买了某种商品之后，给你推荐买这种商品的消费者还会买的其他商品。而训练自动驾驶的AI，则需要采取另外一种机器学习模式：科研人员让AI不再静态地观察汽车驾驶录像，而是处在与环境交互的动态场景中，让AI根据场景实时作出执行动作，并直接考察动作所导致的结果，获得及时反馈。这种机器学习模式被称为强化学习（reinforcement learning）。

1.《信息技术 人工智能 术语》（GB/T 41867-2022）第3.2.37条。

以上是根据全国信息安全标准化技术委员会（"TC260"）发布的《信息安全技术 机器学习算法安全评估规范（征求意见稿）》（"《机器学习算法安全评估规范》"），其中附录B依据训练样本包含的信息以及反馈方式的不同，将机器学习算法分为了监督学习、无监督学习和强化学习三类。

除该等规范之外，我国针对AI监管的立法还存在于《数据安全法》《互联网信息服务深度合成管理规定》《网络信息内容生态治理规定》等法律法规和各项标准中。结合《生成式人工智能服务管理暂行办法》（下称"《AI管理办法》"）提出的要求，我们可以得出初步结论如下：我国针对AI的监管措施，主要可以关注以下三个方面的内容，分别是（1）数据，尤其是训练数据来源及内容等方面的合法性；（2）生成结果，尤其是生成结果中内容的合法性；（3）算法本身的合法性。本文将主要从AI训练相关数据的角度，讨论生成式人工智能中值得关注的合规问题。

/PART 001

何为人工智能的训练相关数据

根据国家标准《信息技术 人工智能 术语》（GB/T 41867-2022）第3.2小节的第3、34、35条相关条款，人工智能训练相关数据包括"用于训练机器学习模型的输入数据样本子集"的训练数据，"用于评估最终机器学习模型性能"的测试数据，以及"用于评估单个或者多个候选机器学习模型性能的数据样本"的验证数据等。

由于《数据安全法》等已生效的法律法规并未对以上三类数据作区分监管，且《互联网信息服务深度合成管理规定》第14条，《AI管理办法》第7条、第19条等条款针对"训练数据"提出额外要求。因此在本文中，我们将主要着眼于"训练数据"，展开讨论当前AI训练相关数据的合规问题，确保AI服务的提供过程和结果的合法性。

关于训练数据，欧洲方面用其来指代"用于通过拟合其可学习参数（包括神经网络的权重）来训练AI系统的数据"[2]，该用法体现在欧洲议会的内部市场委员会和公民自由委员会于2023年5月11日通过的《欧盟人工智能法案》（草案）（"**《欧洲AI法案》**"）中。

/PART 002

现行或已颁布的法律法规对于训练数据的相关要求

1.《生成式人工智能服务管理暂行办法》

2023年7月10日，国家互联网信息办公室（"**国家网信办**"）发布《AI管理办法》，对"研发、利用生成式人工智能产品"的行为作出规定。《AI管理办

2. 'training data' means data used for training an AI system through fitting its learnable parameters, including the weights of a neural network.

法》第7条要求，AI服务提供者应当依法对生成式人工智能产品的预训练数据、优化训练数据来源的合法性负责，并遵守以下具体规定：

（1）使用具有合法来源的数据和基础模型；

（2）涉及知识产权的，不得侵害他人依法享有的知识产权；

（3）涉及个人信息的，应当取得个人同意或者符合法律、行政法规规定的其他情形；

（4）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；

（5）《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律、行政法规的其他有关规定和有关主管部门的相关监管要求。

根据上述第1、5项规定，企业应当遵守《网络安全法》等法律法规，具有合法的数据来源，相对来说这一规定内容存在较为明确的执行标准。但是，关于第2项"不得侵害他人依法享有的知识产权"，这一项的执行对于企业来说存在一定挑战，企业需要主动判断：（1）训练数据是否涉及知识产权（尤其是著作权或商业秘密）；（2）该等知识产权是否存在且合法有效；（3）是否存在该等知识产权利害关系人；（4）相关行为是否侵犯知识产权；（5）行为相关的抗辩条款是否适用等。目前暂未发现国内企业的训练数据涉及侵犯知识产权的报道。

另外，就第4项要求，即数据的"真实、准确、客观、多样"，在适用上可能存在一定的裁量空间，如新闻领域和艺术领域场景对于"数据真实性"的要求就可能略有不同。值得注意的是，鉴于《AI管理办法》将征求意见稿中的"保证"替换成了"增强"，企业可以在确保符合法律法规要求的基础上，参考人工智能使用领域等要素进行灵活调整。

《AI管理办法》第四条还指出，在选择训练数据的过程中，应当采取措施防止出现民族、信仰、国别、地域、性别、年龄、职业、健康等歧视。

2.《互联网信息服务深度合成管理规定》

2022年11月25日，国家网信办等发布了《互联网信息服务深度合成管理规定》，自2023年1月10日起施行。该规定第14条强调，训练数据应当符合个人信息保护的有关规定。

3.《信息安全技术 机器学习算法安全评估规范（征求意见稿）》

2021年7月27日，TC260等发布了《机器学习算法安全评估规范》。该规范对AI训练数据提出的要求主要体现在在以下三个方面：

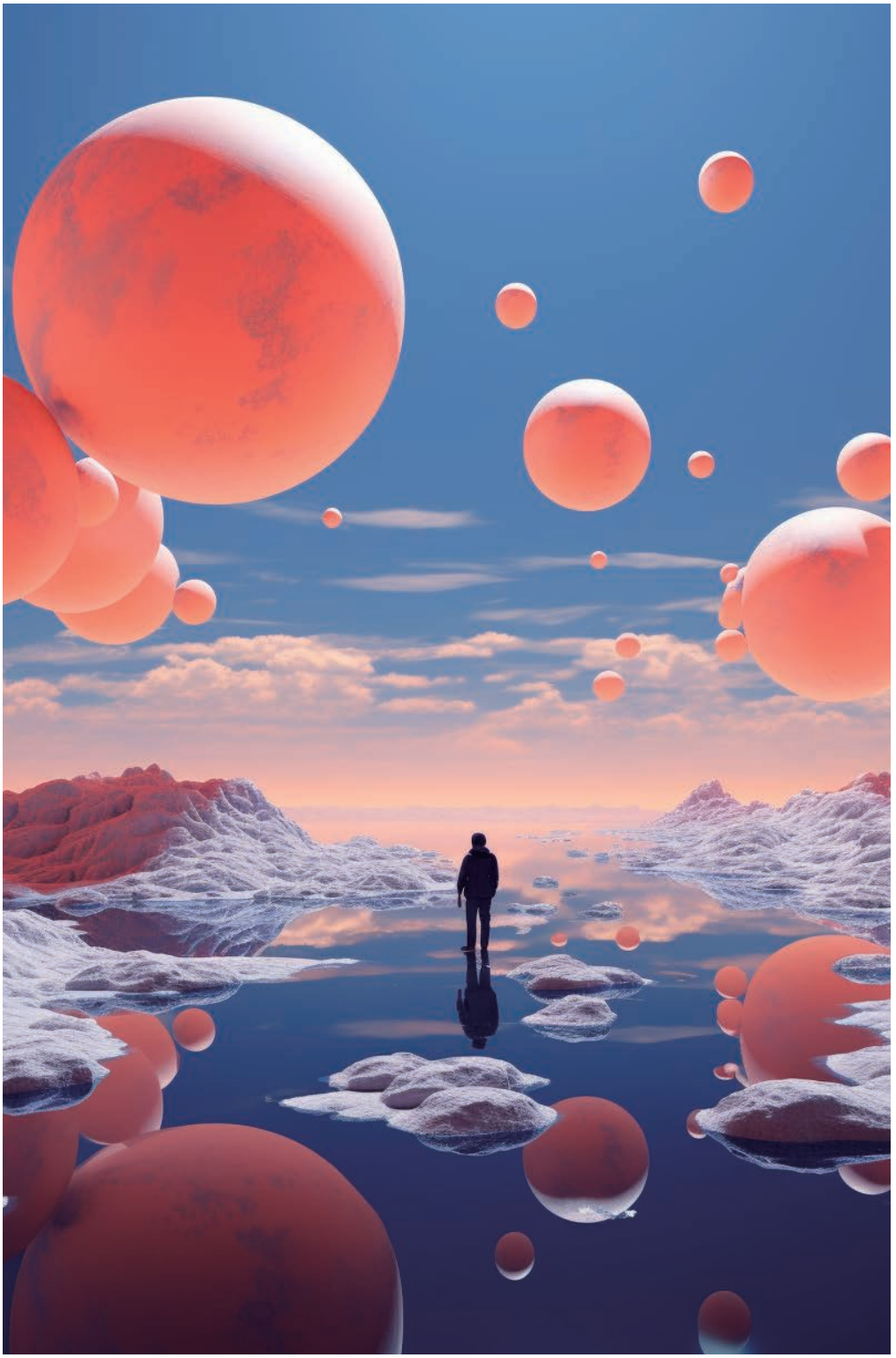
（1）【数据保密和完整】相关组织和个人在开发或运营机器学习算法时，应确保机器学习算法模型、数据、依赖信息的保密性、完整性和可用性，采取措施防范未经授权的访问、篡改、替换或破坏，建立日志，并及时校验数据，使数据的格式和大小等属性处于可用状态（第6.1条）。

（2）【个人信息和隐私合规】相关组织和个人在开发或运营机器学习算法时，应确保处理数据遵守法律和法规要求，保护个人信息和隐私，避免存储、泄漏敏感数据，包括但不限于（1）未经个人同意或法律另有规定，不应使用其个人信息开展机器学习算法相关活动；（2）对个人信息采用必要的数据脱敏措施（第6.1条）。

除上述其他法律法规、国家标准之外，我们目前尚未观察到其他法律规定和公开的行业标准对训练数据及相关事项作出进一步规定。

4.《人工智能安全标准化白皮书（2023版）》

2023年5月29日，TC260发布《人工智能安全标准化白皮书（2023版）》（"**《白皮书》**"），《白皮书》梳理了人工智能技术与应用的发展现状，分析了人工智能面临的安全新风险。《白皮书》指出，网络安全的基本属性包括了AI系统及其相关数据的机密性、完整性、可用性以及针对恶意攻击的防御能力。数据应当具有透明性（用户能够在必要时候获取模型有关信息）、



可解释性（在计算过程中使用的数据、算法、参数和逻辑等对输出结果的影响能够被人类理解）、公平性（不引入偏见和歧视因素）和隐私性（采取隐私增强方案，如最小化数据处理范围、个人信息匿名化处理、数据加密和访问控制等）。

5. 总结

结合以上现行的法律法规以及已颁布的征求意见稿、白皮书等进行分析，训练数据的合规要求可以总结为以下三个方面：

第一，训练数据应当符合网络和数据等**安全合规**方面的要求，如经过数据分类、备份和加密等措施，并存储在境内；

第二，训练数据应当遵循知识产权和个人信息等**权益保护**方面的要求；

第三，训练数据本身应当**可靠透明**，如真实准确、客观中立，具有可解释性和公平性。

/PART 003

境外法律法规对训练数据的要求

（一）联合国方面

联合国教科文组织在2021年发布的《人工智能伦理问题建议书》建议，会员国应当确保人工智能系统的训练数据集：

1.具有透明度和可理解性；

2.不会助长文化、经济或社会不平等和偏见；不会散播虚假信息和错误信息；不会干扰表达自由和信息获取。

（二）欧洲方面

基于《欧洲AI法案》条款第44条达成的共识，《欧洲AI法案》第10条对

高风险AI系统[3]的训练数据集作出了如下规定：

1.应当遵循适当的数据治理和管理实践要求，如对所需数据集的可得性、数量和适用性进行事先评估；审查可能的偏见等。

2.应具有相关性、代表性、无差错和完整性，还应具有适当的统计学意义。

3.应在预期目的要求的范围内，考虑高风险AI系统旨在用于特定地理范围、行为或功能设置的特定特征或元素使用。

4.为了保护他人免受AI系统中的偏见可能导致的歧视，供应商应处理特殊类别的个人数据，以确保对高风险AI系统的偏见进行监测、检测和纠正。

（三）美国方面

《白皮书》中指出，相较于欧盟，美国监管要求少，主要强调安全原则。美国参议院、联邦政府、国防部、白宫等先后发布《算法问责法（草案）》《人工智能应用的监管指南》《人工智能道德原则》《人工智能权利法案蓝图》《国家网络安全战略》等文件，提出风险评估与风险管理方面的原则，鼓励企业将美国的法律法规要求、安全监管原则、主流价值观等置入产品。以生成式人工智能为例，企业一般会在产品设计阶段加入符合安全要求的定制化内容，将其作为重点训练数据。其中，白宫科技政策办公室发布的《人工智能权利法案蓝图》对（训练）数据提出的要求，可以概括为：

1.收集和使用的数据应当仅限于训练或验证机器学习模型，收集和使用行为应当是合法、必要的，尊重个人信息主体权利并符合个人信息主体的预期。

2.用于自动系统开发、评估和部署的数据，应当具有相关性、高质量并适合当前任务。

3.高风险AI系统，根据《法案》的解释性备忘录第5.2.3条，主要有（1）拟用于接受第三方事前合格评定的产品的安全组件的AI系统；（2）附件三中明确列出的，主要涉及基本权利的其他独立的AI系统（附件三所列系统所涉风险已经出现，或在不久的将来有可能出现）。

3.任何在系统开发或评估过程中使用的数据，应当对部署地的社群具有代表性，并经过历史偏见和社会偏见的审查。

（四）小结

纵观联合国、欧洲和美国，可以看出AI系统的主要研发国家和地区对训练数据的要求主要集中于：

1.数据来源和使用行为合法，符合信息主体预期并尊重其权利；

2.数据应当具有相关性、代表性、无差错和完整性，且应当尽量避免偏见；

3.有部分国家和地区还规定了训练数据的实效性，如数据质量、统计意义等纳入监管要求。

/PART 004
对企业的人工智能训练数据的建议

我们理解，企业通常通过以下两种方式收集训练数据：（1）直接收集，主要通过软硬件产品进行搜集，如互联网应用程序、软件开发工具包（SDK）、Cookies、互联网智能家居，甚至采取网络爬虫等技术手段等收集数据；（2）间接收集，即其他数据处理者自愿提供，如由其他数据处理者授权访问、查阅、下载、传输等。我们认为，无论采取直接或间接收集的方式获取训练数据，企业均应当关注数据来源、数据性质、使用目的与方式，关注其是否符合针对训练数据的合规要求。

值得注意的是，当设置为AI爬取训练数据的网络爬虫时，应当格外关注合规要求，及相应网站的反爬虫声明等。据澎湃新闻报道[4]，加州一家律师事务所表示，OpenAI从互联网上秘密抓取了约3000亿字的内容，其中包括书

4.https://www.thepaper.cn/newsDetail_forward_23679287，诉讼文书原文https://assets.bwbx.io/documents/users/iqjWHBFdfxIU/rIZH4FXwShJE/v0。

籍、文章、网站和帖子，甚至还包括未经同意的个人信息。基于该等情况，该律所提出了30亿美元潜在损失的赔偿要求。

因此，建议企业收集训练数据时，按照数据来源对训练数据进行分类分级，确保《网络安全法》《个人信息保护法》相关规定得到充分落实。针对爬取数据，应当进一步确认数据来源是否设置了反爬取声明或协议，以及数据本身是否涉及数据来源相关企业的商业秘密等。如数据涉及个人信息的，企业需关注相关的信息是否获得个人的明确授权，授权内容与企业意图采取的处理方式是否匹配等；在使用数据前对其进行分类分级，并采取相应的保密措施。企业在准备训练数据时，可以根据《互联网信息服务管理办法》《网络信息内容生态治理规定》等规定，对数据是否含有不良信息进行识别和审查，优化训练数据以符合主流价值观，推进人工智能技术依法合理有效利用。同时不容忽略的是，企业准备训练数据时应当根据AI的使用场景，追求"真实、准确、客观、多样"，来有效提高AI在合规基础上的实用性。必要时，可以引入外部律所针对训练数据的合规性做出专业评估和指导，以满足相关法律法规的要求。

此外，我国与欧美等地对训练数据的要求存在一定差异，如果企业未来计划拓展境外市场，或者有可能向来自该等国家和地区的人员提供服务的，企业还需关注自身训练数据是否符合该等国家和地区的法律法规，以确保运营安全。必要时可引入具有涉外服务经验的律所进行专业评估，出具专业法律意见，为企业的境外市场运营保驾护航。

（崔晓霞对本文亦有贡献）

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

透视AIGC产品的生命周期

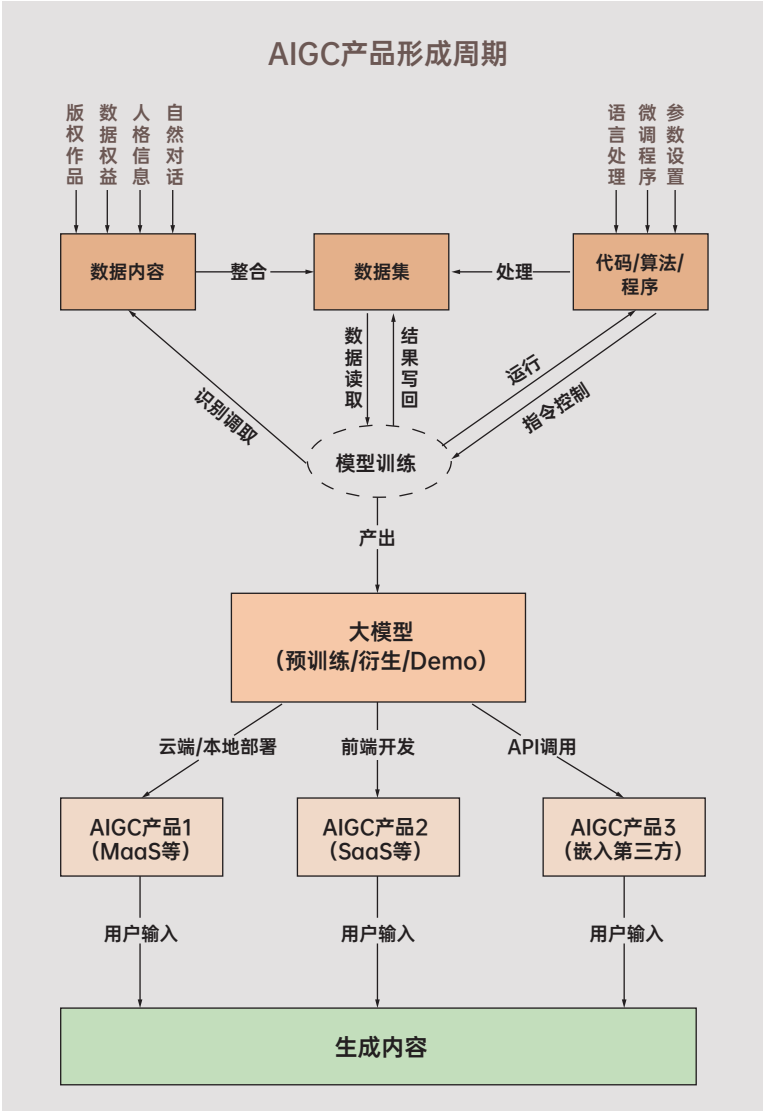
——数据与代码的授权合规

ARTICLE BY 王飞

本文将从AIGC产品生命周期剖析AIGC产品在0到1过程中数据与代码方面的知识产权合规风险，为AIGC产品合规提供有益借鉴。

/PART 001

AIGC产品的诞生



相关权利

数据内容知识产权
数据权益
个人信息、隐私
汇编作品著作权
程序、代码著作权
算法专利

著作权
商业秘密

软件著作权
外观设计知识产权

生成内容著作权
（争议）

如前图所展示，AIGC产品从0开始到训练模型、到形成大模型、到AIGC产品、最终到收到用户指令输出生成内容，整个生命周期将历经6个重要风险合规要点。

1、数据内容。数据内容是用于模型学习的具体信息的总称，例如用于自然语言处理训练的人类对话语料文本、用于人工智能生成图像训练的画作、照片等。除少数不承载任何个性化信息或智力成果的内容外，绝大多数数据内容受到法律保护，涉及著作权、隐私权、肖像权、个人信息权益、数据权益以及其他财产权益。

2、数据集。数据集是对数据内容进行归集后形成的用于模型学习的大体量信息集合。目前主流的大模型训练所用的数据集一般包括十亿以上文本单位（Token），少数可达万亿文本单位级别。[1]ChatGPT的基础模型GPT-3训练所用的数据集中，有60%来自于开放数据网站Common Crawl在2016-2019年间随机爬取的全球互联网信息网页快照形成的数据集，其原初体量高达45TB，经过筛选和清洗后，最终用于训练的高质量数据集仍有570GB。[2]高质量的数据集可能作为汇编作品受到著作权法保护，数据集提供者享有相应著作权。

3、代码、算法与程序。既包括搭建大模型的程序、命令时撰写的基本代码和应用的算法，也包括在搭建模型功能时所调用的各类调试工具、模型、程序等。例如在搭建模型时常用的指令分词的分词器（Tokenizer）程序，典型代表是谷歌开源的SentencePiece模型[3]；再如对大模型参数增加

1.GenAI, Meta. (2023). Llama 2: Open Foundation and Fine-Tuned Chat Models. P.20. https://scontent.fhkg3-1.fna.fbcdn.net/v/t39.2365-6/10000000_662098952474184_2584067087619170692_n.pdf?_nc_cat=105&ccb=1-7&_nc_sid=3c67a6&_nc_ohc=RYfzDCymkuYAX-Sq5_b&_nc_ht=scontent.fhkg3-1.fna&oh=00_AfDu7ph4Nn-a3xUmrpt5rpG9TZqsp1sJ8VeyWtFFLw8YqQ&oe=64C25B7F

2.Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., & Dhariwal, P. et al. (2020). Language Models are Few-Shot Learners. P.8. Retrieved 17 July 2023, from https://arxiv.org/abs/2005.14165

3.Google. (2023). GitHub - google/sentencepiece: Unsupervised text tokenizer for Neural Network-based text generation. Retrieved 17 July 2023, from https://github.com/google/sentencepiece

附加层实现针对特定领域微调的PEFT技术，典型代表是微软开源的LoRA模型[4]。前述模型在大模型的自然语言处理环节和调试内容产出环节能够发挥重要作用。这些代码、模型可能作为计算机程序而受到著作权法保护，算法可能作为数据处理或信息分析的方法发明专利受到专利法保护，非公开内容亦可能作为商业秘密受到保护。

4、大模型。大模型是通过特定算法架构对数据集进行深度学习后形成的独立的程序，通过大规模参数赋予其回应指令、解决问题的技能，生成符合指令的结果。一般认为，一个模型完成预训练后即可被归入大模型范畴，此后经过微调（Fine-tune）、优化和修改，再衍生出各类产出内容精确度更高、更符合使用者需要的各类衍生大模型或Demo版本。大模型是AIGC产品形成流程中的核心权利节点，目前广为人知的大模型包括OpenAI公司的GPT系列模型（其中，ChatGPT就是基于GPT-3.5微调而成），Meta公司的LlaMA系列模型，清华大学的GLM-130B模型，以及百川公司最新发布的Baichuan-17B模型等。大模型可能作为作品受到著作权法保护，亦可能作为商业秘密受到保护。

5、AIGC产品。基于大模型形成的，经过前端代码开发或通过API接口而形成的，能够供用户直接使用的程序、软件、云端的MaaS（模型即服务）或SaaS（软件即服务）。例如，OpenAI基于GPT-3.5模型开发出的ChatGPT聊天程序在chat.openai.com上向用户开放，用户只要注册并同意使用条款后就可以在该页面与ChatGPT进行对话；同时其模型亦通过开放API接入在线笔记应用Notion形成Notion AI[5]，使Notion的用户可以使用GPT

4.Microsoft. (2023). GitHub - microsoft/LoRA: Code for loralib, an implementation of "LoRA: Low-Rank Adaptation of Large Language Models." Retrieved 17 July 2023, from https://github.com/microsoft/LoRA

5.Notion AI Supplementary Terms. (2023). Retrieved 24 July 2023, from https://www.notion.so/Notion-AI-Supplementary-Terms-fa9034c8b5a04818a6baf3eac2adddbb

模型优化笔记内容。AIGC产品作为一个完整的软件或程序，在大模型本体之外仍有大量功能开发等代码，其完整形态可作为计算机软件受到著作权法保护，相应的用户界面设计、显示屏幕面板、界面视图等还可能作为外观设计专利获得保护。

6、生成内容。用户向AIGC产品输入自然语言指令后，由AIGC产品通过模型处理输出的生成内容。目前主流AIGC产品的生成内容已涵盖各题材文本、图像、画作、代码、视频等。有些大模型尚未开发出AIGC产品，但有一定计算机技能的工程师也可以在特定配置环境中调用其源代码生成内容。AIGC生成内容是否构成著作权法意义上的"作品"是国内外讨论激烈的话题。本文笔者认为，AIGC生成内容有可能作为作品获得著作权法保护（具体讨论可参见笔者《ChatGPT的著作权规制》一文），并将以此为基点展开后文讨论。简言之，AIGC生成内容可能产生著作权等知识产权权益，值得严肃对待。

/PART 002

数据内容的授权合规

1、数据内容的风险

当前，部分数据内容的收集欠缺相关制度规范，处于野蛮生长阶段，大量数据内容在权利人不知情的情况下被机械爬取、复制进入数据集。此类爬取和复制行为处于民事侵权乃至刑事犯罪的灰色地带，以至于最终形成的数据集在来源合规方面也存在较高风险。例如，某些数据集可能完整复制了某作者在某网站上发布的文章、某艺术家发布并禁止转载的画作、某社交软件用户在交谈中透露的个人信息等。这些数据内容的访问和复制往往未经权利人许可，存在侵犯著作权、个人信息权益、肖像权之虞。

2023年1月至2月，某国际大型版权图片提供商在美国和英国先后对AI绘

画公司Stability AI提起诉讼，认为被告在训练旗下的图片生成模型Stable Diffusion时擅自复制了120万张版权照片及其题注文字、元数据用于训练。[6]7月，三位美国作家对OpenAI[7]和Meta[8]分别提起集体诉讼，认为两公司擅自复制了其作品，用于GPT大模型和LlaMA大模型训练。上述诉讼同时附带商标权侵权、违约、不正当竞争等指控，原告提出要求被告停止侵权、删改所有涉及权利作品的数据内容及生成内容等诉求，在赔偿方面分别主张应得利益损失或惩罚性赔偿等。

此外，数据收集的个人信息，不仅包括以文本或数字形式出现的用户在互联网平台上公开的个人信息，还可能涉及个人照片、声音、手写字迹等高度敏感的生物识别信息。这些信息往往在权利人不知情的情况下被爬取和使用，且欠缺通知和退出机制。2023年6月28日，美国16名原告对OpenAI提起集体诉讼，认为其从社交网站抓取用户的个人信息、视频、音频等用于模型训练违反了美国《联邦电子通讯隐私法案》、伊利诺伊州《生物识别信息隐私法》等。[9]除此之外，还有一种常见的数据内容收集方式，即从商业机构处购买用户个人信息。除非用户对此明确知情且同意，否则该行为仍有违背个人信息处理的正当、必要原则的合规风险。

机械抓取的数据内容中可能存在虚假、非法、偏见、歧视、危害身心健康等负面内容。如果未经清洗和筛选就将上述内容用于模型训练，很可能对模型及AIGC产品的最终效果产生不利影响。例如，2016年3月发生的某AI机器人"被教坏"事件：人工智能机器人"Tay"通过与用户在社交媒体上互动进行实时

6.Complaint. Getty Images v. Stability AI. (2023). https://aboutblaw.com/6DW. Retrieved 24 July 2023.
7.Class Action Complaint. S. Silverman et al. v. Open AI. (2023).
https://www.documentcloud.org/documents/23869693-silverman-openai-complaint
8.Class Action Complaint. R. Kadrey et al. v. Meta Platforms. (2023).
https://llmlitigation.com/pdf/03417/kadrey-meta-complaint.pdf
9.Class Action Complaint. P. M. et al. v. Open AI. (2023).
https://clarksonlawfirm.com/wp-content/uploads/2023/06/0001.-2023.06.28-OpenAI-Complaint.pdf

机器学习并作出回应，将真人用户输入的信息自动纳入数据集并进行学习反馈，最终导致其在上线十几个小时后就开始对外发表种族主义、极端主义言论，最终被下架。

2、数据内容的授权合规路径

(1)权利自持：生成原创数据内容开展训练

通过生成原创的数据内容训练模型，可以将模型开发中的数据合规风险降低，从根源上把控模型及后续产品合规。尽管以目前的模型体量来看，全原创数据难以满足大模型预训练所需的数据量级，但仍有望用于训练小体量模型或模型微调。例如，2023年4月12日Databricks发布的Dolly 2.0模型在基于开源模型EleutherAI pythia系列上进行人工指令微调，该微调所用的数据集全部来自该公司5000名员工在2023年3月至4月原创撰写的15000个高质量文本素材。[10]这一例子展示了通过原创数据避免权利瑕疵的可能。

此外，高质量的原创数据也可以对数据内容的整体价值起到积极效果，产生小数据集撬动大模型的效果。随着算法、模型的优化和数据资源开发枯竭，未来的模型训练也可能走向数据体量浓缩化，通过权利自持的原创数据内容进行模型训练将具备更高的实用价值。

(2)授权使用：向数据内容权利人获取授权

模型开发者可以先从权利人处获得合法授权，再将相关信息和内容用于模型训练。拥有社交平台类产品的企业可以将授权条款写入用户协议，从而获取相应授权，以便使用用户在平台上传、发布的信息训练模型。

应当注意的是，采取此种方式获得授权应遵循格式条款有关规范。考虑到

10.Free Dolly: Introducing the World's First Truly Open Instruction-Tuned LLM. (2023). Retrieved 18 July 2023, from https://www.databricks.com/blog/2023/04/12/dolly-first-open-commercially-viable-instruction-tuned-llm

用户在平台上传的内容既可能涉及著作权、肖像权，又可能涉及个人信息乃至敏感信息（例如人脸特征），因此这类授权条款的设计需要审慎考量。从合规角度考虑，有必要根据《民法典》第496条第2款之规定，采取合理方式履行提示或者说明义务，例如设置强制弹窗勾选程序等。此外，如涉及敏感信息收集，还应适当限缩相关授权权限范围。最高人民法院"法释〔2021〕15号"司法解释明确规定，通过格式条款获得授权收集人脸信息的，不得为无限期、不可撤销、可任意转授权等授权形式，否则该条款无效。在获取授权后，开发者处理个人信息同样应遵循个人信息相关保护规范，有必要配置告知同意程序和通知删除功能。

对于适用于特定场景的定向模型，开发者可与数据持有人合作，由数据持有人提供数据集并对数据内容作权利无瑕疵保证。2017年，某大型互联网公司与生物科技公司Adaptive合作开发基于血液检测的免疫系统信息解码模型，其训练所用的数据全部来自于Adaptive提供的生物和医疗信息，Adaptive并做出权利无瑕疵保证。[11]

(3)数据清洗：对数据内容进行清洗和优化

《办法》第七条明确，生成式人工智能服务提供者在训练数据处理时应采取有效措施增强训练数据的真实性、准确性、客观性、多样性。在数据筛选和清洗的过程中，相关服务的提供者可以增加合规处理环节，去除违法、虚假、歧视内容，对敏感个人信息进行脱敏处理，对文字作品、视听作品等著作权客体采取摘要、引用等方式纳入数据集。上述做法既有助于提高数据内容质量，也可有效降低合规风险。目前，国际领先的大公司已开始重视在大模型训练中加强数据内容合规处理。例如，Meta公司在2023年7月19日发布大模型

11.Lee, P. (2018). M* and Adaptive Biotechnologies announce partnership using AI to decode immune system; diagnose, treat disease - The Official M* Blog. Retrieved 18 July 2023, from https://blogs.microsoft.com/blog/2018/01/04/microsoft-adaptive-biotechnologies-announce-partnership-using-ai-decode-immune-system-diagnose-treat-disease/

LLaMA-2时即表示，该公司在预训练前清洗并排除了数据集中已知的、来自含有大量个人隐私信息的网站的数据，以强化LLaMA-2模型的安全性，避免侵犯他人信息和隐私安全。[12]

/PART 003

数据集/代码的授权合规

1、数据集/代码开源的风险

得益于信息技术社区的开放共享传统，目前市面上存在大量的开源模型、代码、工具，为大模型和AIGC产品研发日新月异的进展铺路赋能。也正因如此，许多开发者选择基于既有的数据集、代码、预训练模型等建立或调试自己的大模型。在调用他人成果时，应格外注意是否属于他人成果的许可条款授权范围，避免因违反许可协议而产生违约或侵权责任。

非开源的工作成果往往要求后续开发者单独签署许可协议以获取授权，合同明文规定使许可内容和范围更加清晰明确，也方便当事人知晓并遵守合同中的重要事项；而在开源社区中，许可协议往往仅以"License.txt"为名悄然列于项目文件列表，使用者如果不慎忽视就将面临重大法律风险。开源许可协议种类繁多，模板化的开源协议集群例如MIT，BSD，GPL，Apache，CC等；每个集群下又有若干协议版本，例如GPL集群下的AGPL，LGPL，CC集群下的CC BY-SA等。开发者也可能设置个性化的开源协议，代表如为GPT提供绝大多数自然语言训练语料的Common Crawl数据集、Meta公司开源的LLaMA-2

12.GenAI, Meta. (2023). Llama 2: Open Foundation and Fine-Tuned Chat Models. P.20. https://scontent.fhkg3-1.fna.fbcdn.net/v/t39.2365-6/10000000_662098952474184_2584067087619170692_n.pdf?_nc_cat=105&ccb=1-7&_nc_sid=3c67a6&_nc_ohc=RYfzDCymkuYAX-Sq5_b&_nc_ht=scontent.fhkg3-1.fna&oh=00_AfDu7ph4Nn-a3xUmrpt5rpG9TZqsp1sJ8VeyWtFFLw8YqQ&oe=64C25B7F

大模型都发布了个性化的许可协议。开源协议的形态或显繁杂，但万变不离其宗。模型开发者在查询开源协议/许可证时应重点注意以下风险要点。

(1)是否可修改/改编

是否可修改/改编决定着能否调用该数据集和代码用于个人模型开发。在调用数据集和代码来建立或调试个人模型时，往往会调整原数据集或代码的应用范围、参数等，但如果许可证中不包括可修改，意味着前述调整都将超出许可范围；同理，如果许可证中不包括可改编（Derivative），意味着他人无权调用该数据集或代码建立或调试个人模型，也不能从原模型基础上转换、调整、混合出Demo版本。

(2)是否可商用

是否可获得商用许可决定了某一开源工具能否超越科研领域而用于营利目的。尽管开源，许多数据集、模型、工具禁止商业使用，例如清华大学的GLM-130B模型、斯坦福大学的Alpaca-7B模型。部分版本的开源协议也直接排除了商业使用许可，例如CC集群下带有"NC"（Non-commercial）标志的协议均意味着非商用。如果开发者调用了不可商用的许可协议下的数据集和代码，由此产生的模型就仅能以非营利的方式使用。

(3)是否传染与强制开源

开源协议的"传染"是指，调用了此类协议的开源数据集或代码形成的工作成果必须以相同开源协议对外分发，且首先意味着强制开源。典型的传染性开源协议如GPL、Mozilla，以及CC协议集群中带有"SA"（Share Alike）标志的协议。传染性使得部分商业开发者对开源软件望而却步，例如谷歌规定不得将AGPL开源协议的软件用于公司工作，防止谷歌的代码被强制开源。[13]除强制

13.AGPL Policy | Google Open Source. (2023). Retrieved 23 July 2023, from https://opensource.google/documentation/reference/using/agpl-policy

开源之外，不同开源协议的传染性亦有所差异，例如GPLv2协议要求后续程序以相同协议分发且所有条款不得增删变更，而GPLv3则只要求核心条款不得变更，但允许使用者增删其他条款，例如附加免责条款。

(4)是否存在非竞争条款

基于开放共享的理念，主流开源协议大多不含非竞争条款，但在个性化的开源协议中，非竞争条款却愈发常见。例如，OpenAI的使用条款中要求使用者不得将OpenAI的输出结果用于开发与其有竞争关系的模型；[14]再如，LLaMA-2虽是开源、可商用的模型，但其许可协议同样限制不得将LLaMA-2的代码、生成内容及其他输出结果用于改进任何其他大模型（LLaMA-2的衍生模型除外）。[15]

2、数据集/代码的授权合规路径

(1)查看许可协议并遵照执行

为尽可能控制法律风险，建议模型开发者在调用他人数据集、代码、模型、工具前首先查看许可协议，明确可获得的授权范围。在调用多个数据集、代码的情形下，还应注意不同许可协议的兼容问题，例如GPLv2协议与Apache 2.0协议不兼容，因为Apache 2.0缺少前者要求的某些专利中止与侵害保护条款；[16]再如，两个不同的传染性许可协议也可能导致适用冲突。开发者可以在开源协议组织者的官网查询许可协议的兼容情况。

(2)采取技术手段隔绝传染性

开发者如果既想调用具有传染性的开源协议下的代码或数据集，又希望规

14.Open AI Terms of use. (2023). Retrieved 23 July 2023, from https://openai.com/policies/terms-of-use
15.Facebookresearch. (2023). llama/LICENSE at main • facebookresearch/llama. Retrieved 23 July 2023, from https://github.com/facebookresearch/llama/blob/main/LICENSE
16.Various Licenses and Comments about Them - GNU Project - Free Software Foundation. (2023). Retrieved 23 July 2023, from https://www.gnu.org/licenses/license-list.html#GPLCompatibleLicenses

避掉传染性要求，则可以考虑采取一定开发或技术手段平衡商业需要并降低法律风险。

在开发端，可考虑采取"净室方法"（Cleanroom approach）隔绝传染性，该方法最早由IBM公司提出，基本内容是安排两组团队，团队一负责代码编写但不阅读有传染性的源代码，团队二阅读源代码、研究其原理并向团队一说明，但不参与任何开发和代码编写工作。前述过程同时安排律师进行监督。[17] 这一方法基于著作权法仅保护表达而不保护思想或方法的基本原理，通过隔绝开发团队与传染性代码的接触以免受传染性许可协议约束。在技术端，也可以考虑采用封装代码、管道通信等技术方式对调用代码和自建代码实现静态或动态隔离，以降低因代码传染而被迫开源的法律风险。

(3)前溯检查他人代码调用来源

在调用他人代码时，也应前溯检查他人代码是否调用了外部URL和数据包，以避免因他人代码调用侵权资料而受牵连。例如2019年"阿里巴巴公司与荣耀公司'掌上小说'APP著作权侵权纠纷案"中，被告调用了一段开源代码，该代码被解码后发现其实际调用了侵权数据内容，法院最终判决调用该开源代码的被告同样构成著作权侵权。

17.R. C. Linger and H. D. Mills, "A case study in cleanroom software engineering: the IBM COBOL Structuring Facility," *Proceedings COMPSAC 88: The Twelfth Annual International Computer Software & Applications Conference, Chicago,* IL, USA, 1988, pp. 10-17, doi: 10.1109/CMPSAC.1988.17141

/PART 004

小结

在生成式人工智能迅猛发展的当下，风险与挑战也正悄然而生。在这场长期且颠覆性的技术竞争中，只有于方兴未艾之际预判并控制风险，才能行稳致远。本文对AIGC产品形成过程的数据内容、数据集和代码的相关授权合规提供借鉴。

（尼婧瑶对本文亦有贡献）

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

AIGC数据跨境的法律监管和合规路径

ARTICLE BY 蔡荣伟 斯响俊 杨杰

在生成式人工智能（Generative AI，又称AI Generated Content，下称"AIGC"）技术的发展和应用过程中，相关法律监管问题一直备受各国政府关注。例如，AIGC相关技术和硬件的出口管制问题、预训练数据的数据合规问题、AIGC生成物的可版权性和权利归属问题、电信监管及行业监管问题以及科技伦理问题等。相关法律问题直接关系到各国AIGC技术的健康发展和广泛应用。

国家互联网信息办公室（"国家网信办"）于2023年7月10日发布了《生成式人工智能服务管理暂行办法》（"《AIGC服务暂行办法》"），该办法已于2023年8月15日正式实施。该办法是我国针对AIGC技术服务专门出台的一个部门规章，表明了我国对于AIGC技术服务的发展及其合规监管的重视。

《AIGC服务暂行办法》中明确了，AIGC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。"[1]可见，AIGC是基于模型及相关技

1.《生成式人工智能服务管理暂行办法》第22条。

术而搭建的，而要训练出成熟可用的模型需要海量的数据。因此，数据合规问题是AIGC技术发展和应用过程中不可避免的重要课题。《AIGC服务暂行办法》亦特别指出，AIGC服务提供者应当依法开展预训练、优化训练等训练数据处理活动，使用合法来源的数据及基础模型，遵守《中华人民共和国网络安全法》（"《网安法》"）、《中华人民共和国数据安全法》（"《数安法》"）、《中华人民共和国个人信息保护法》（"《个保法》"）等法律、行政法规的有关规定和有关主管部门的相关监管要求。[2]

AIGC数据合规是一个复杂的议题，需要讨论的问题颇多。本文将主要围绕其中一个重要问题——AIGC的数据跨境合规问题进行探讨，以期为可能涉及数据跨境的AIGC开发者和服务提供者提供参考。

2.《生成式人工智能服务管理暂行办法》第7条。

/PART 001

AIGC数据出境主要场景分析

从目前实践来看，在AIGC技术开发过程中，主要存在以下数据出境场景：

1.跨境调用算力导致数据出境

AIGC底层模型的开发不仅需要海量的数据推动大模型"涌现"功能的出现，还需要强大的算力作为支撑。但是，目前国内AIGC产业链的基础设施层（主要包括芯片技术和云计算平台）仍待完善。与此同时，鉴于高性能AI芯片对国家竞争力的重大影响，美国也相继出台了一系列政策措施，以限制美国高性能AI芯片的对华出口。

因此，目前境内AIGC开发者可能存在算力不足的问题。在该种情况下，不少AIGC开发者在探讨跨境调用境外算力的可能性。如AIGC开发者跨境调用境外算力，则其采集的训练数据将会被传输至境外进行训练，相关训练数据和搭建后的模型亦可能被存储至境外数据中心，从而引发数据跨境相关风险。

2.调用境外算法模型导致数据出境

考虑到目前一些境外的AIGC算法和模型更为成熟和先进，故在实践中，AIGC开发者采用的更为直接的方式是通过境外模型平台(Model as a Service)或其他方式直接调用境外的算法模型来训练自己的定制化模型。在这一过程中，境内的AIGC开发者需要将其采集的相关行业数据、业务数据等数据传输至境外用以模型训练，从而引发数据跨境相关风险。

/PART 002

数据出境法律监管和合规路径

自《网安法》《数安法》和《个保法》相继发布和实施后，我国数据出境监管的基本框架初步搭建。此后，《数据出境安全评估办法》以及配套的《数据出境安全评估申报指南（第一版）》《个人信息保护认证实施规则》《网络安全标准实践指南——个人信息跨境处理活动安全认证规范（第二版）》《个人信息出境标准合同办法》等法规和标准相继颁布，进一步完善了我国数据出境监管体系，并促进了相关监管措施的落地和实施。

目前，我国数据出境监管体系主要围绕对个人信息和重要数据出境的监管。在现有监管体系下，个人信息和重要数据出境的合规路径主要如下：

1.个人信息出境

(1)个人信息出境的前置合规

a.履行告知义务

根据《个保法》相关规定，个人信息处理者在向境外提供个人信息前，应当履行以下告知义务：

i.一般告知义务——个人信息处理者在处理个人信息前，应当以显著方式、清晰易懂的语言真实、准确、完整地向个人告知下列事项：(i)个人信息处理者的名称或者姓名和联系方式；(ii)个人信息的处理目的、处理方式，处理的个人信息种类、保存期限；(iii)个人行使《个保法》规定权利的方式和程序；以及(iv) 法律、行政法规规定应当告知的其他事项。[3]

ii.特殊告知义务——个人信息处理者向境外提供个人信息的，应当向个人

3.《个人信息保护法法》第17条。

告知境外接收方的 (i) 名称或者姓名、联系方式；(ii) 处理目的、处理方式、个人信息的种类；以及 (iii) 个人向境外接收方行使《个保法》规定权利的方式和程序等事项。[4]

b.取得个人信息主体的单独同意

根据《个保法》相关规定，个人信息处理者在向境外提供个人信息前，应当取得个人信息主体的单独同意。[5]

c.开展个人信息保护影响评估

根据《个保法》相关规定，个人信息处理者向境外提供个人信息的，应当事前进行个人信息保护影响评估，并对处理情况进行记录，且个人信息保护影响评估报告和处理情况记录应当至少保存三年。[6]

(2)个人信息出境的三种路径

在完成以上所述的个人信息出境的前置合规程序后，个人信息处理者应视其所处行业及其处理个人信息的数量等因素选择其数据出境的合规路径。根据《个保法》及其配套法规，个人信息出境主要有以下三种路径：

a.向国家网信办申报数据出境安全评估

根据《数据出境安全评估办法》相关规定，个人信息出境符合以下情形之一的，应当通过所在地省级网信部门向国家网信办申报数据出境安全评估[7]：

i.关键信息基础设施运营者向境外提供个人信息；

ii.处理100万人以上个人信息的数据处理者向境外提供个人信息；以及

iii.自上年1月1日起累计向境外提供10万人个人信息或者1万人敏感个人信息的数据处理者向境外提供个人信息。

4.《个人信息保护法法》第39条。
5.《个人信息保护法法》第39条。
6.《个人信息保护法法》第55、56条。
7.《数据出境安全评估办法》第4条。

b.与境外接收方订立标准合同并备案

根据《个人信息出境标准合同办法》相关规定，个人信息处理者为非关键信息基础设施运营者，且其处理的个人信息的数量未达到申报数据出境安全评估标准的，可依法通过与境外接收方订立标准合同的方式向境外提供个人信息。[8]

标准合同应当严格按照《个人信息出境标准合同办法》附件的范本订立。个人信息处理者应当在标准合同生效之日起10个工作日内向所在地省级网信部门备案。备案时应当提交(i)标准合同，以及(ii)个人信息保护影响评估报告。[9]

c.进行个人信息保护认证

如个人信息处理者为非关键信息基础设施运营者，且其处理的个人信息的数量未达到申报数据出境安全评估标准的，其可根据《个保法》相关规定，通过专业机构进行个人信息保护认证。

根据《个人信息保护认证实施规则》相关规定，个人信息保护认证的认证模式为：技术验证+现场审核+事后监督。认证机构根据认证委托资料、技术验证报告、现场审核报告和其他相关资料信息进行综合评价，作出认证决定。对符合认证要求的，颁发认证证书。认证证书有效期为3年。在有效期内，通过认证机构的获证后监督，保持认证证书的有效性。[10]

2.重要数据出境

根据《数据出境安全评估办法》相关规定，数据处理者向境外提供重要数据的，应当通过所在地省级网信部门向国家网信办申报数据出境安全评估。[11]

根据《数安法》相关规定，国家数据安全工作协调机制统筹协调有关部门

8.《个人信息出境标准合同办法》第4条。
9.《个人信息出境标准合同办法》第7条。
10.《个人信息保护认证实施规则》第3、4.4、5.1.1条。
11.《数据出境安全评估办法》第4条。

制定重要数据目录，各地区、各部门应当按照数据分类分级保护制度，确定本地区、本部门以及相关行业、领域的重要数据具体目录。[12]但截止目前，仅有少数个别行业制定了重要数据识别的行业标准，如《汽车数据安全管理若干规定（试行）》及《YD/T 3867-2021基础电信企业重要数据识别指南》，分别界定了汽车行业及基础电信行业的重要数据。但是，大多数行业的重要数据目录仍待进一步明确。

3.国家发布数据出境新规征求意见稿，数据出境合规程序或可简化

2023年9月28日，国家网信办发布了《规范和促进数据跨境流动规定（征求意见稿）》（"**《数据跨境流动征求意见稿》**"）。如果该征求意见稿正式实施，或可简化部分企业数据出境的合规程序，降低其数据出境过程中所需履行的合规成本，具体如下：

（1）个人信息出境

对于个人信息出境场景，《数据跨境流动征求意见稿》主要从以下方面豁免了相关企业的数据出境合规义务：

α.设置出境白名单，明确特定场景下，无需要申报数据出境安全评估、订立个人信息出境标准合同、通过个人信息保护认证[13]（以下合称**"数据出境合规程序"**）：

i.不是在境内收集产生的个人信息向境外提供的，可豁免数据出境合规程序。例如，AIGC企业从境外收集的训练数据，在境内处理后传输出境的情况，可无需再履行任何数据出境合规程序。

ii.为订立、履行个人作为一方当事人的合同所必需，如跨境购物、跨境汇款、机票酒店预订、签证办理等，必须向境外提供个人信息的。

12.《数据安全法》第21条。

13.《规范和促进数据跨境流动规定（征求意见稿）》第三、四条。

iii.按照依法制定的劳动规章制度和依法签订的集体合同实施人力资源管理，必须向境外提供内部员工个人信息的。

iv.紧急情况下为保护自然人的生命健康和财产安全等，必须向境外提供个人信息的。

b.不再将个人信息处理总量作为考量因素，而是以预计一年内向境外提供的个人信息数量作为考量标准。具体而言[14]：

i.预计一年内向境外提供不满1万人个人信息的，无需履行任何数据出境合规程序。本条实质降低了仅涉少量个人信息出境企业的合规负担。

ii.预计一年内向境外提供1万人以上、不满100万人个人信息的，无需申报数据出境安全评估，仅需订立并备案个人信息出境标准合同或者通过个人信息保护认证的。

iii.预计一年内向境外提供100万人以上个人信息的，应当申报数据出境安全评估。

此外，《数据跨境流动征求意见稿》并未区分一般个人信息和敏感个人信息的合规门槛，敏感个人信息出境的更为严格的计算标准或将被弱化。

c.允许自贸区自行制定负面清单，对于负面清单外数据出境，可以豁免数据出境合规程序。[15]与白名单相比，"负面清单"仅保留了对清单内数据的监管，无疑是采取了更为宽松的监管策略，为自贸区建立了先行先试的数据跨境流通专有通道。

虽然《数据跨境流动征求意见稿》在特定场景下豁免了部分企业的数据出境合规程序，但值得注意的是，《数据跨境流动征求意见稿》并未免除《个保法》等相关法律法规下对于个人信息保护的基本要求。例如，个人信息处理者需履行的"告知 - 同意"义务；《个保法》第55条项下的进行个人信息保护影

14.《规范和促进数据跨境流动规定（征求意见稿）》第五、六条。
15.《规范和促进数据跨境流动规定（征求意见稿）》第七条。

响评估的义务，同样未被豁免。

（2）重要数据出境

《数据跨境流动征求意见稿》中规定，未被相关部门、地区告知或者公开发布为重要数据的，数据处理者不需要作为重要数据申报数据出境安全评估。[16]对于属于重要数据目录尚未制定行业的企业，这一规定无疑可以解决他们的困惑。相关企业无需再主动自查出境数据是否属于重要数据，而是可等待被动告知，而后再采取相关合规措施。

/PART 003

对AIGC数据出境的合规建议

针对AIGC数据跨境的主要场景（详见本文第二部分），结合我国现行数据出境法律监管和合规体系（详见本文第三部分），我们总结了以下AIGC数据出境相关合规建议，供可能涉及数据出境的AIGC开发者及服务提供者提供参考。

1.注意识别训练数据中的重要数据及个人信息

在调用境外算力或模型进行训练的过程中，可能涉及训练数据的出境。因此，在训练数据采集、清洗和标注的过程中，应注意识别相关训练数据中是否包含重要数据及个人信息。尤其是在模型定制场景下，可能涉及大量特定行业和业务场景数据的出境，如果相关行业属于关乎国家安全、公共利益的重要或敏感行业，则应特别注意相关行业数据是否会落入重要数据的范畴。

如经识别，相关训练数据中确实包含重要数据及个人信息，则应根据国家相关规定，积极采取合规措施；如无法确定是否包含，则建议与相关监管机构

16.《规范和促进数据跨境流动规定（征求意见稿）》第二条。

及时沟通或咨询专业机构进行确定，以减少和规避相关法律风险。

2.根据数据来源判定合规风险及责任

目前，AIGC的训练数据主要来源于(i)在自身业务中直接采集或生成的数据，如银行直接在业务中获取的客户个人信息，以及形成的相关业务数据（"**直采数据**"）；(ii)通过互联网获取的数据，如通过爬虫爬取的相关互联网数据（"**互联网数据**"）；以及(iii)通过数据交易方式向专门的数据提供商购买数据（"**交易数据**"）。

根据训练数据的来源不同，企业应注意采取不同的措施履行合规义务：

(1)对于直采数据，如其中包含个人信息，则企业在该等训练数据的采集阶段，即应注意对采集对象履行充分告知义务，并取得采集对象的单独同意。

(2)对于互联网数据，其数据来源一般较为复杂，难以追溯，本身即可能存在合规风险。因此，应当尤其注意对此类数据的清理和处理，尽量确保该类数据不包含任何个人信息和重要数据。

(3)对于交易数据，企业则应注意在与数据提供商的合同中，明确数据将会被跨境传输，并将数据出境的合规义务转移给数据提供方。例如，如交易数据中包含个人信息的，数据提供方应保证其已经履行了个人信息出境合规义务，包括已向个人信息主体履行了充分告知义务，并取得个人的单独同意。

3.及时履行数据出境合规程序

AIGC企业应注意识别相关出境数据中是否包含重要数据和个人信息，并根据出境数据的类型以及个人信息的数量及时判定其是否需要履行相应数据出境合规程序（如数据出境安全评估、进行个人信息出境标准合同等）。

虽然网信办发布了《数据跨境流动征求意见稿》，或可豁免部分企业履行数据出境合规程序的义务。但是，该征求意见稿尚处于向社会公开征求意见阶段，各企业仍应按照国家现行数据出境的监管要求，积极采取各项合规措施，

以防范和应对数据出境相关风险。

AIGC技术的发展日新月异，相关监管措施也必定会日益完善。本文仅从AIGC数据出境这一议题切入，进行了理论层面的探讨。但是，AIGC在实践中遇到的法律问题必定更为复杂。因此，AIGC开发者和服务提供者应密切关注其技术开发及技术服务提供所在国家和地区的相关法律监管措施及趋势，以减少和规避相关法律风险。

蔡荣伟
高级顾问
公司业务部
上海办公室
+86 21 6061 3175
roncai@zhonglun.com

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

A

CHAPTER

04

AIGC监管

C

G

全景透视生成式AI的法律挑战(三):

监管合规挑战与应对

ARTICLE BY 陈际红 陈煜娘

除数据合规与知识产权问题外，在中国法语境下，AIGC技术的应用还可能会面临互联网信息服务和信息内容治理、算法合规、增值电信监管、科技伦理等多层面的行政监管。本篇将对在中国境内提供AIGC服务所适用的主要监管框架进行梳理，并就我国AIGC监管的几个重点问题进行分析。

/PART 001

框架梳理：境内提供AIGC服务的主要监管框架

（一）算法监管

AIGC技术的底层逻辑是算法和模型，自2021年起，主管部门以"每年一部"的频次，相继出台了数部关于算法监管的规定。2021年12月31日，国家网信办联合四部门发布《互联网信息服务算法推荐管理规定》（以下简称"**《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》（以下简称"**《深度合成管理规定》**"）正式出台；2023年7月，《生成式人工智能服务管理暂行办法》正式发布，其中亦对AIGC涉及的算法提出合规要求。至此，我国涉及AIGC算法监管的法律框架正式形成，AIGC服务提供者应当依法履行算法相关监管要求。

（二）互联网信息服务及信息内容监管

基于AIGC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AIGC服务可能构成"提供互联网信息服务"[1]，并需承担相应信息内容监管责任(具体而言，根据《暂行办法》，应承担内容生产者责任)。

以2011年修订的《互联网信息服务管理办法》为核心，主管部门陆续出台了《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，AIGC服务提供者应结合自身业务模式（例如是否利用AIGC技术从事"经营性互联网文化活动""互联网视

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。

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。

（三）增值电信监管

目前，不少AIGC服务提供者将AIGC技术嵌入其他垂直领域进行应用，以为用户提供更好的产品体验。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），可能构成开展增值电信业务，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。AIGC服务提供者应结合AIGC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。

（四）科技伦理监管

相较于征求意见稿，《生成式人工智能服务管理暂行办法》将《科学技术进步法》作为上位法，突出AIGC技术的科技伦理因素；《算法推荐管理规定》《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。具体而言，依据相关规定[2]，科技部等部门于2023年10月发布《科技伦理审查办法（试行）》，该办法落地后，将成为科技伦理审查方面普遍适用的规则，AIGC服务提供者应遵循此等规则。

除上述针对算法、互联网信息服务及信息内容、增值电信、科技伦理等监管规则外，针对AIGC技术本身，《生成式人工智能服务管理暂行办法》亦提出了一系列针对性合规要求，例如使用具有合法来源的数据和基

2.具体依据为《科学技术进步法》《关于加强科技伦理治理的意见》。

础模型、制定数据标注规则[3]等。具体合规要求请见《AIGC合规义务清单》。

/PART 002

焦点探析：境内AIGC监管的几个重点问题

问题一：在境内提供AIGC服务需要履行哪些备案、评估程序？

目前，大部分AIGC技术会嵌入某网站或APP并面向公众提供服务，结合《互联网信息服务管理办法》相关规定，依法构成在境内提供互联网信息服务，需根据其是否为"经营性"而取得相应ICP许可/备案，与此同时，根据《计算机信息网络国际联网安全保护管理办法》相关规定，凡是接入互联网并向境内提供服务，均需开展公安联网备案。此等ICP许可/备案与公安联网备案系面向境内提供互联网信息服务的一般性要求，针对AIGC技术本身，所涉及的特定备案及评估要求主要包括：

- **算法备案：**具有舆论属性或者社会动员能力的算法推荐服务提供者应向网信部门开展备案（《算法推荐规定》第24条）。"舆论属性或者社会动员能力"的范围相当宽泛[4]，基于AIGC技术之"内容生成"的产品样态且通常会嵌入其他垂直领域的网络产品/服务，我们理解其落入算法备案的可能性较高。2023年9月1月，国家网信办发布第二批境内深度合成服务算法备案信息，文心大模型、京东言犀大模型等多家大型模型已然在列。
- **安全评估：**互联网信息服务提供者在涉及特定情形[5]时，需自行开展

3.《生成式人工智能服务管理暂行办法》第7、8条。

4.根据《舆论属性安全评估规定》第2条，"舆论属性或社会动员能力"的情形包括：（一）开办论坛、博客、微博客、聊天室、通讯群组、公众账号、短视频、网络直播、信息分享、小程序等信息服务或者附设相应功能；（二）开办提供公众舆论表达渠道或者具有发动社会公众从事特定活动能力的其他互联网信息服务。

安全评估，并向网信部门和公安机关提交安全评估报告（《舆论属性安全评估规定》第3条、第7条）。基于AIGC这一崭新的技术样态，我们理解，其触发"具有舆论属性或社会动员能力的信息服务上线，或者信息服务增设相关功能的"或"使用新技术新应用...，导致舆论属性或者社会动员能力发生重大变化的"的可能性较高，应依法履行安全评估义务。

－"双新评估"：结合监管实践和行业理解，提供AIGC服务可能还需开展"互联网新技术新业务安全评估"（又称"**双新评估**"），可供参考的标准为《YD/T 3169-2020 互联网新技术新业务安全评估指南》《YD/T 3738-2020 互联网新技术新业务安全评估实施要求》。从实践来看，通过"双新评估"的难度较高，但作为目前我国针对新类型网络产品或服务上线的重要监管手段，建议相关AIGC服务提供者仍应按照相关指南及要求，积极开展互联网新技术新业务安全评估，拥抱监管。

问题二：AIGC服务提供者应对生成内容承担什么责任？

《暂行办法》第九条规定，AIGC服务提供者应当依法承担网络信息内容生产者责任，履行网络信息安全义务。"网络信息内容生产者"的概念源自《网络信息内容生态治理规定》，指"制作、复制、发布网络信息内容的组织或者个人"，与之相区分的概念是"网络信息内容服务平台"，即"提供网络信息内容传播服务的网络信息服务提供者"[6]。**基于此，在信息内容监管层面，监管部门要求AIGC服务提供者直接承担内容生产责任，**

5.根据《舆论属性安全评估规定》第3条，应开展安全评估的情形包括：
（一）具有舆论属性或社会动员能力的信息服务上线，或者信息服务增设相关功能的；（二）使用新技术新应用，使信息服务的功能属性、技术实现方式、基础资源配置等发生重大变更，导致舆论属性或者社会动员能力发生重大变化的；（三）用户规模显著增加，导致信息服务的舆论属性或者社会动员能力发生重大变化的；（四）发生违法有害信息传播扩散，表明已有安全措施难以有效防控网络安全风险的；（五）地市级以上网信部门或者公安机关书面通知需要进行安全评估的其他情形。

6.《网络信息内容生态治理规定》第41条。

而非平台责任。

具体而言，《暂行办法》第4、14条针对信息内容提出了两个方面的监管要求：

- 提供AIGC服务不得生成法律、行政法规禁止的内容[7]（即"违法信息"），关于"违法信息"，可参照《网络信息内容生态治理规定》第6条进行识别和判断；

- AIGC服务提供者如发现违法内容，应当及时采取停止生成、停止传输、消除等处置措施，采取模型优化训练等措施进行整改，并向有关主管部门报告。

除上述"不得生成的**违法信息**"的要求外，《网络信息内容生态治理规定》第7条还进一步对内容生产者提出了针对"不良信息"的要求，即"应采取措施，防范和抵制制作、复制、发布**不良信息**"，我们理解，此等要求亦适用于AIGC服务提供者。

问题三：在境内提供AIGC服务需要哪些具备电信牌照？

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"，基于此，将AIGC技术嵌入网站或APP并向公众提供内容生成服务本身构成"互联网信息服务"，进而需根据是否为"经营性"而依法相应取得ICP许可/备案[8]。

除ICP许可/备案外，我们理解，仅基于纯粹的AIGC技术面向公众提供服务，直接触发其他具体增值电信许可牌照的可能性较低；但目前AIGC技术在各垂直领域的应用已越发广泛，AIGC服务提供者应结合具体应用类型（例如提供短信回复服务、电商服务），判断整体上是否需取得

7.《暂行办法》第4条。

8.《互联网信息服务管理办法》第4条，国家对经营性互联网信息服务实行许可制度；对非经营性互联网信息服务实行备案制度。未取得许可或者未履行备案手续的，不得从事互联网信息服务。

相应增值电信许可证，例如SP证、EDI证等。

问题四：如何进行AIGC生成内容标识？

《暂行办法》第12条规定，AIGC服务提供者应当按照《深度合成管理规定》对图片、视频等生成内容进行标识。具体而言，《深度合成管理规定》第16、17条规定了两类标识要求：

- **一般性标识。**对于使用一般性深度合成服务生成的内容，应当采取技术措施添加不影响用户使用的标识，并保存日志信息。
- **显著标识。**对于特殊的深度合成服务[9]（例如智能对话、智能写作等模拟自然人进行文本的生成或者编辑服务）生成的内容，可能导致公众混淆或者误认的，应当在生成或者编辑的信息内容的合理位置、区域进行显著标识。

我们理解，基于AIGC之"生成式"的技术原理，结合目前ChatGPT等智能对话、智能写作的典型业态，AIGC服务提供者应当对基于AIGC技术生成的内容进行**显著标识**，且此等标识应当达到避免公众混淆或误认的效果。具体实操层面，可以参照信安标委于2023年8月发布的《网络安全标准实践指南——生成式人工智能服务内容标识方法（征求意见稿）》（以下简称"**《标识方法》**"），例如，在显示区域下方或使用者输入信息区域下方持续显示提示文字，或在显示区域的背景添加包含提示文字的显式水印标识；由人工智能生成图片、视频时，应采用在画面中添加提示文字的方式进行标识，提示文字宜处于画面的四角，所占面积应不低于画面的0.3%或文字高度不低于20像素，提示文字应至少包含"由人工智能生

9.根据《深度合成管理规定》第17条，此等特殊的深度合成服务类型包括：
（一）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；（二）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；（三）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；（四）沉浸式拟真场景等生成或者编辑服务；（五）其他具有生成或者显著改变信息内容功能的服务。

成"或"由 AI 生成"等信息等。

/PART 003

结语

ChatGPT一经问世，即以最快速度突破亿级用户增长，AIGC的快速发展使得科幻小说中的种种想象即将成为触手可及的现实，是"辅助工具""人机协作"还是"取代竞争"，关于人工智能与人类的关系引发了关于技术创新的又一轮讨论。正如同阿尔文•托夫勒在《第三次浪潮》中所言，面对科学技术发展所带来的剧变和不安，我们需要讨论的是一个模式和这份希望。尽管存在AIGC技术应用可能面临数据合规、知识产权与监管层面的诸多法律挑战，我们仍对AIGC在国内发展的前景充满信心，也对中国式治理实践充满期待。

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

万紫千红待新雷：
《生成式人工智能服务管理暂行办法》立法解读

ARTICLE BY 蔡鹏 肖莆羚令

生成式人工智能是一种利用算法与数据自动生成新内容的技术，在各行各业，特别是科技、文化、教育、娱乐和新闻等多个领域，具有广泛的应用和发展空间。然而，生成式人工智能也面临着数据安全、知识产权、伦理道德等方面的挑战和风险。

2023年7月10日，国家网信办会同六部委共同发布《生成式人工智能服务管理暂行办法》（"《办法》"）。《办法》充分吸收了此前各界对于征求意见稿的反馈意见，明确坚持发展和安全并重、促进创新和依法治理相结合的原则。

在未来的一段相当长的时间里，生成式人工智能势必成为科技产业和数据合规领域的热门议题。本文试图探析《办法》背后的立法理念和监管路径，以期一窥生成式人工智能未来的发展方向和监管趋势。

/PART 001

立法之谨：合规义务的精准调整

在2023年四月发布的《生成式人工智能服务管理办法（征求意见稿）》（下称"**征求意见稿**"）中，涉及生成式人工智能内容安全的要求占据了较大的篇幅。这些要求被以法定义务的形式施加给企业，却没有充分考虑企业在落地过程中面临的现实困境，甚至可能挑战生成式人工智能的技术逻辑。这一点在征求意见稿发布后受到了广泛的讨论。

《办法》对征求意见稿中的生成式人工智能服务提供者（"**提供者**"）的合规义务进行了优化调整，为企业在谋求发展与满足合规义务之间预留了一定缓冲地带，充分体现了立法的谦抑性。

一方面，《办法》较之征求意见稿减轻了企业的合规负担，包括将企业对训练数据真实性、准确性的控制义务由征求意见稿中命令式的100%保证义务降格为尽职性的努力承诺，大大降低了企业因难以核实全部训练数据真实性、准确性而被迫违法的风险（第七条）；考虑到了提供者对使用者生成内容的有限控制能力，删除了禁止提供者生成歧视性内容的规定。

另一方面，《办法》为提供者创设了一定的"发挥空间"。提供者可以通过服务协议方式教育、监督生成式人工智能服务使用者的合法使用，控制并合理转嫁合规风险(第九条)。同时，征求意见稿中关于提供者一旦发现网络炒作、恶意发帖跟评、制造垃圾邮件、编写恶意软件必须暂停或终止服务的规定被修改为可以依法依约灵活采取警示、限制功能，暂停或终止服务等多种措施，给予提供者一定的自由裁量权(第十四条)；征求意见稿中关于提供者发现违规生成内容需在3个月内优化训练模型的规定也一并删除，尊重提供者在提升算法模型性能和提高内容管理方面的主观能动性。

此外，全国信息安全标准化技术委员会于2023年10月发布《生成式人工智能服务安全基本要求》（征求意见稿），从技术角度对生成式人工智能预料安全、模型安全、安全措施以及安全评估等方面内容作出进一步说明，作为对《办法》中各项合规义务的落地标尺。

/PART 002

立法之勉：生成式人工智能产业的支持政策

《办法》在原则性条款和具体规定中，都体现了国家对生成式人工智能产业发展的鼓励和支持的态度，既为产业创新提供了政策导向和法律保障，又为产业监管提供了科学合理和平衡适度的框架。

首先，《办法》在征求意见稿基础上新增原则性条款，明确提出国家坚持发展和安全并重、促进创新和依法治理相结合的双重原则，为生成式人工智能产业监管划定了基调（第三条）。其次，《办法》将《科学技术进步法》作为其上位法，强调其对推动人工智能服务科技进步的核心理念，蕴含着在高效、协同、开放的国家创新体系下，充分发挥市场配置创新资源的立法导向（第一条）。

再次，《办法》针对生成式人工智能服务提出了多元化、多层次的鼓励方案。在技术层面上，鼓励算法、框架、芯片、配套软件平台等产业链环节、产品层级的创新；在基础设施建设上，鼓励各级政府和企业共同参与生成式人工智能基础设施和公共训练数据资源平台的建设，促进算力资源协同共享（第六条）；在市场参与度上，支持行业组织、企业、教育和科研机构、公共文化机构、有关专业机构等各类角色在生成式人工智能技术创新、数据资源建设、转化应用、风险防范等方面的协作配合（第五条）；在资源投入上，推动利用公共数据进行算法训练；在产业扶植上，鼓励企业采购安全、可信赖的芯片、软件、工具、算力和数据资源。

这些规定将在未来数年的生成式人工智能的"军备竞赛"中发挥效用，为企业和国家在全球化的竞争格局中提供了安全保障和稳定支持。

/PART 003

立法之巧：行业监管的特性和路径

与征求意见稿相比，《办法》更加突出行业监管的特性。第十六条规定，网信、发展改革、教育、科技、工业和信息化、公安、广播电视、新闻出版等部门，依据各自职责依法加强对生成式人工智能服务的管理。国家有关主管部门针对生成式人工智能技术特点及其在有关行业和领域的服务应用，完善与创新发展相适应的科学监管方式，制定相应的分类分级监管规则或者指引。这或许意味着，此后对于生成式人工智能的监管将以行业为线索，逐步呈现差异化和针对性。

这种行业导向的监管模式符合生成式人工智能的强技术属性，也有利于各行业根据具体需求为生成式人工智能服务制定更合理的监管标准和措施，既能充分发挥生成式人工智能服务的经济价值，又能有效防范合规风险。例如，对于在新闻业中应用的生成式人工智能，如何防止人工智能生成、传播虚假新闻可能是重要的监管目标；而对于金融行业的生成式人工智能，如何保证其对用户背景和画像的客观性和公正性，以及如何应对系统遭受攻击时对业务连续性的影响则更为紧迫。

通过行业为线索的监管方式，各部门能够更加精准地理解和管理生成式人工智能服务，为不同行业制定具体的监管措施和指导方针，共同推动我国生成式人工智能服务行业的健康发展。

/PART 004

立法之智：全球监管的核心抓手

在全球范围内，各国对于人工智能的监管也在积极探索和实践中。从目前已经出台或正在制定的人工智能监管法规来看，算法透明度和分类分级是两大共同关注的核心议题。

算法透明度是指算法的设计、训练、优化、运行等过程能够向外界公开或解释其原理、逻辑、数据、结果等信息，以便进行有效的监督和评估。算法透明度对于监管和社会的可持续发展至关重要。透明的算法可以使用户和相关利益相关方更好地理解系统的决策依据，从而增加对人工智能系统的信任和接受度，同时可以帮助监管机构审查和评估人工智能系统的合规性和公平性，确保其不会产生歧视性、偏见或不当的行为。此外，透明的算法能够促进人工智能技术的创新和进步，通过学习和纠正错误，提高系统的性能和效果。《办法》新增提供者需"基于服务类型特点，采取有效措施，提升生成式人工智能服务的透明度，提高生成内容的准确性和可靠性"就是最直观的体现（第四条）。此外，《办法》援引《互联网信息服务算法推荐管理规定》对算法备案的要求，也是一种确保算法透明度的方式（第十七条）。

分类分级是指根据人工智能技术及其应用场景对可能产生的风险和影响进行不同程度地划分，并采取相应强度的监管措施。通过分类分级的监管框架，监管机构可以根据人工智能系统的应用领域、风险程度和技术成熟度，制定相应的监管要求和措施。这有助于确保监管更加精准和针对性，避免一刀切的监管模式。同时，分类分级的监管框架也可以促进创新和发展，因为根据不同级别的监管要求，企业和研究机构可以更好地规划和管理技术研发，降低合规成本和风险。《办法》已经明确提出了人工智能分类分级的要求，这种监管方式也与《数据安全法》对于数据分类分级保护的手段一致，不排除后续行业主管部门或可就人工智能分类分级和数

据分类分级两方面要求形成相关联的分类分级规则（第十六条）。

在中国以外的其他地区，算法透明度和分类分级的监管思路也在各国立法中有所呈现和实践，例如：

- 欧盟《人工智能法案（草案）》将人工智能系统分为不可接受的风险、高风险、有限风险和低/轻微风险四类，针对不同级别的人工智能系统设置了不同的合规要求，并要求高风险人工智能系统需在专有数据库中备案；
- 美国《数据隐私和保护法案（草案）》要求，相关实体在应用算法时，需向联邦贸易委员会提交算法影响评估和算法设计评价，对这些文件进行备案，并应要求向国会提交前述文件；
- 加拿大发布的涉及人工智能监管的法律框架草案中明确，将对具有重大影响的人工智能系统实施更为严格的监管规则。

随着全球生成式人工智能进入黄金发展阶段，相关监管立法也预期呈现井喷式增长，各国对生成式人工智能的监管思路的探索以及异同都是值得持续关注的问题。

/PART 005
结语

生成式人工智能作为一项跨时代的技术，必将对人类社会的演进产生无可估量的影响。生成式人工智能服务不仅拥有广阔应用前景和巨大发展潜力，同时亦面临着诸多挑战和风险。

在2023年11月1日至2日首届人工智能峰会期间，中国、英国、美国、欧盟等二十八个国家和地区的政府代表联合签署了《布莱切利宣言》。《布莱切利宣言》重申了"以人为本、可信和负责任"的人工智能发展理念，并呼吁国际社会加强合作以应对由人工智能引发的安全风险。同

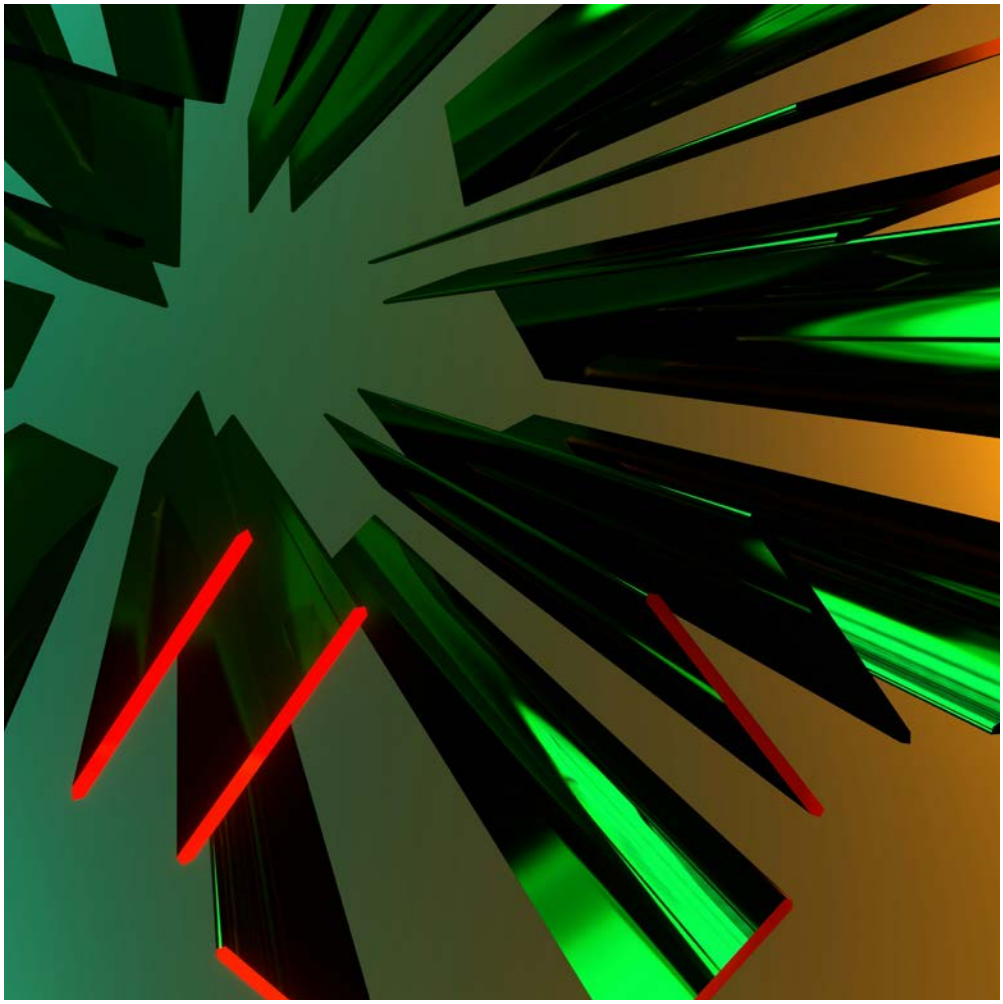
时，它也倡导各国科研机构在人工智能的设计、开发和使用过程中，始终关注人类的福祉，并确保人工智能技术能为人类带来切实的利益。

《办法》作为我国首次对生成式人工智能服务进行专门规范的法规，在立法理念、监管路径、核心问题等方面充分体现了包容审慎、兼容谦抑的姿态。《办法》的出台不仅为我国生成式人工智能服务行业的健康发展和规范应用提供了法律依据和制度保障，也为全球人工智能监管提供了有益的借鉴和参考。

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

跨越AIGC产品合规上市之路(一):

算法备案



ARTICLE BY 蔡鹏 肖莆羚令 陈雨婕

算法备案作为算法治理体系的重要监管内容，是算法透明度要求的落地方式之一，旨在保护用户权益，维护产品安全和信息安全。自《互联网信息服务算法推荐管理规定》（下称《算法推荐管理规定》）、《互联网信息服务深度合成管理规定》（下称《深度合成管理规定》）提出算法备案要求以来，中央网信办已连续发布了四批境内互联网信息服务算法备案清单和首批境内深度合成服务算法备案清单。

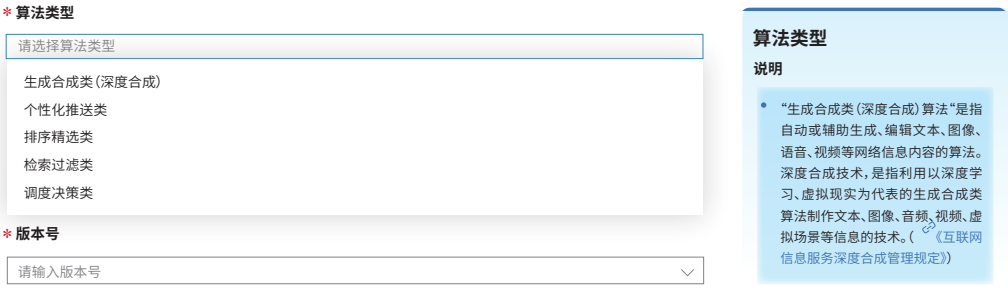
此外，各大应用商店也开始加强对App完成算法备案情况的审核。根据《深度合成管理规定》第十三条，如AIGC产品未按照相关法律法规履行算法备案的，应用商店有权采取不予上架、警示、暂停服务或者下架等处置措施。鉴于此，算法备案义务已成为AIGC产品合规上市的必由之路。

/PART 001

"深度合成算法"与"生成合成类算法"的区别

在《深度合成管理规定》发布之初，业界曾对"深度合成技术"与"生成合成类技术"的关系展开讨论。根据《深度合成管理规定》规定，"深度合成技术，是指利用深度学习、虚拟现实等生成合成类算法制作文本、图像、音频、视频、虚拟场景等网络信息的技术"。从文意本身分析，"深度合成技术"是"生成合成类算法"的利用形态之一。

实践中，"互联网信息服务算法备案系统"显示，"生成合成类算法"与"深度合成算法"被列为一类，即"**生成合成类（深度合成）算法**"。因此，即使在技术层面深度合成技术与生成合成技术的关系或存争议，但就算法备案实操而言，企业履行算法备案义务以及被要求提供的备案信息并不因生成合成类算法与深度合成算法的区别而有所区分。



/PART 002

算法备案实务介绍

1、备案主体

与《算法推荐管理规定》所监管的其他几类算法不同，深度合成算法是唯一一项在备案环节需要区分备案主体身份的算法。其备案角色包括服务提供者和技术支持者，且两类主体对同一算法的备案义务相互独立。

对于企业而言，判定其是否应当履行备案义务时，可参照以下线索：

- **如企业针对同一款算法存在前述两种角色，应当分别完成作为深度合成服务提供者以及深度合成服务技术支持者的备案；**
- **如企业仅为服务提供者，不对外提供深度合成服务技术支持，则其仅需要完成作为深度合成服务提供者的备案；**
- **对于从技术供应商处采购深度合成技术并利用该技术向终端用户提供深度合成服务的企业而言，无论该深度合成技术供应商是否已完成相关备案，其也需以深度合成服务提供者的身份履行单独的备案义务。**

对于集团公司而言，同一个算法可能由集团的多个下属公司共享。在此情形下，集团公司可以选择有权控制算法的公司作为备案主体。但需要注意的是，此情形下算法备案主体可能与应用该算法的AIGC产品的ICP主体不一致。企业需要在主体信息中提交关于算法备案与ICP主体不一致的情况说明，例如二者之间存在关联关系等。但根据我们的实践经验，算法备案与ICP主体不一致的情形往往会因流程上的说明与核实，导致审核进度的延缓。因此对于希望迅速通过算法备案的企业而言，选择与ICP主体一致的公司作为备案主体可能更为有利。

2、备案范围备案程序

结合《<互联网信息服务深度合成管理规定>备案填报指南》以及我们的项目实践，深度合成算法备案的流程包括三个步骤：

- **第一步：填报主体信息**，填报完成后需等待后台工作人员审核通过方可继续填报算法信息和产品及功能信息。
- **第二步：填报算法信息**，包括算法基础属性信息、算法详细属性信息（详见下图）。

算法基础属性填报内容

* 算法类型

请选择算法类型

* 算法名称

请选择算法名称，如算法类型-编号

* 上线时间

请选择日期

版本号

请输入版本号

* 应用领域

请选择应用领域

* 算法安全自评估报告

下载模版 上传
支持pdf(不超过20MB)

上传文件

* 拟公示内容

下载模版 上传
支持pdf(不超过20MB)

上传文件

* 算法简介：

请输入算法简介，限制200字

* 使用场景：

请选择使用场景

算法数据

* 输入数据模态：

请选择输入数据模态

* 输入的人物特征是否包含生物特征： ○ 是 ○ 否

* 输入的人物特征是否包含身份信息： ○ 是 ○ 否

* 输出数据模态：

请选择输出数据模态

* 输出文件格式：

请选择输出文件格式

* 输出数据模态：

请输入文件大小

* 是否支持批量输出： ○ 是 ○ 否

算法模型

训练数据来源(选填)

开源数据集&来源：

数据集名称

描述具体来源

+ 添加

自建数据集&来源：

数据集名称

描述具体来源

+ 添加

合作数据集&来源：

数据集名称

描述具体来源

+ 添加

* 是否包括境外数据： ○ 是 ○ 否

* 产生方式：

请选择产生方式

* 是否涉及个人信息： ○ 是 ○ 否

* 使用的哪种生成合成算法：

请选择使用的哪种生成合成算法

* 算法硬件要求：

请输入算法硬件要求

* 算法性能：

请输入算法性能

* 算法计算方式：

请选择算法计算方式

算法策略

* 是否对训练数据进行预处理(模型策略)： ○ 是 ○ 否

* 是否对用户输入数据进行预处理(输入策略)： ○ 是 ○ 否

* 是否对产出结果进行后处理(输出策略)： ○ 是 ○ 否

算法风险与防范机制

* 是否对自身服务制作的生成合成内容添加隐式标识： ○ 是 ○ 否

* 是否对生成合成的内容进行显著标识： ○ 是 ○ 否

* 是否具备提醒用户对生成合成的内容进行显著标识的能力： ○ 是 ○ 否

* 用户数据安全保障机制：

请选择用户数据安全保障机制

* 对生成合成虚假信息的辟谣机制：

请选择对生成合成虚假信息的辟谣机制

* 对生成合成的不良内容发现处置机制：

请选择对生成合成的不良内容发现处置机制

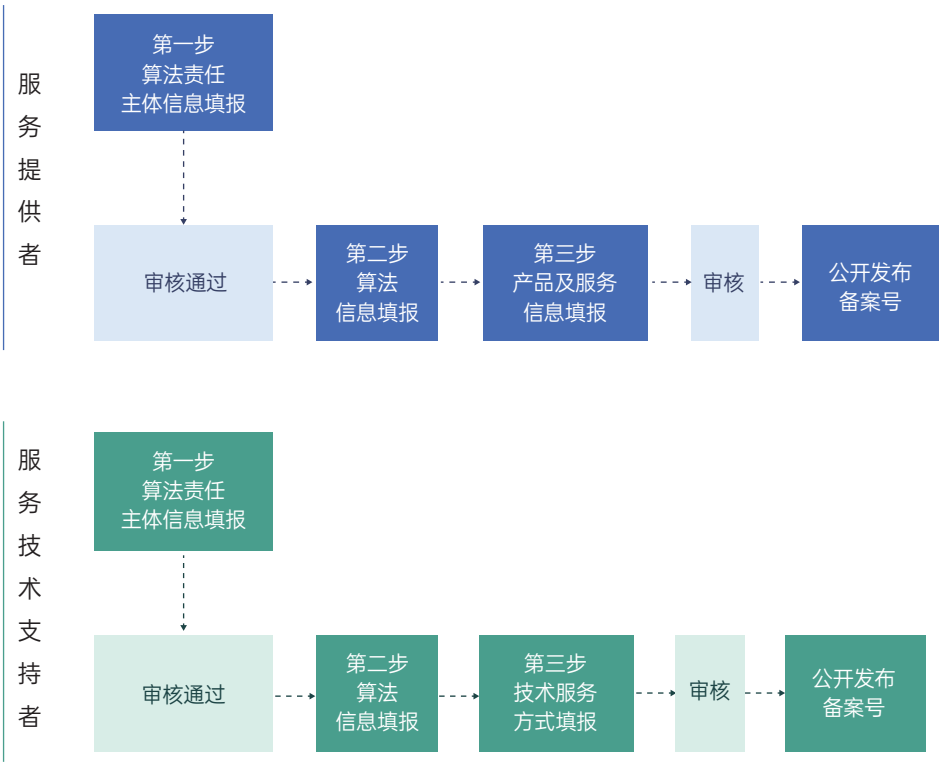
* 风险防范机制说明：

请输入风险防范机制说明，限制200字

算法详细属性信息

- **第三步：关联产品及功能信息或填报技术服务方式**，第二步和第三步需一并递交审核。

深度合成服务提供者和技术支持者在算法备案过程中的差异主要在于第三步填报内容不同，即深度合成服务提供者需填报关联产品及功能信息，而深度合成服务技术支持者需填报技术服务方式。具体流程详见下图：



在具体填报的过程中，鉴于填报内容较多，通常情况下需要一定时间准备材料，因此我们建议企业首先根据"互联网信息服务算法备案系统"梳理需要填报的内容，在确认填报的内容准确无误后再行上传至系统中，以免因系统不稳定或者未保存而需重复操作。

3、备案材料

在算法备案材料中，除了主体信息、算法基本信息等实时填写的基本内容之外，还需单独准备以下几个附件：

(1)落实算法安全主体责任基本情况

在主体信息填报（第一步）中，企业需要上传"落实算法安全主体责任基本情况"表，其主要内容包括企业设置的算法安全专职机构以及制定的算法安全管理制度。算法安全管理制度应当至少包括算法安全自评估制度、算法安全监测制度、算法违法违规处置制度、算法安全事件应急处置制度、科技伦理审查制度等。

(2)算法安全自评估报告

在算法信息填报（第二步）中，企业需要提交算法安全自评估报告，其主要包括算法情况（算法流程、数据、模型和干预策略）、服务情况、风险研判、风险防控情况、安全评估结论等内容。对比服务提供者与技术支持者的算法安全自评估报告模板，服务提供者需提供更多关于内容生态治理、结果标识、辟谣机制，以及用户权益保障方面的说明。

(3)拟公示内容

根据《算法推荐管理规定》第十六条，算法推荐服务提供者应当以适当方式公示算法推荐服务的基本原理、目的意图和主要运行机制等。在算法信息填报（第二步）中，企业即需要提交算法拟公示内容，主要包括算法基本原理、算法运行机制、算法应用场景、算法目的意图，以供监管机构审核。鉴于目前监管机构并未公布关于算法基本原理、算法运行机制等内容的公示维度，建议企业可参考行业目前的实践做法，把握公示内容的颗粒度，以确保在不泄露算法技术商业秘密的前提下履行算法公示的法律义务。

前述材料是整个算法备案中的关键审核材料，其内容不仅涉及算法基本原理、算法属性等算法技术本身，还涉及算法安全管理制度、组织架

构、风险防控措施等算法合规体系构建情况。因此，建议需完成算法备案的企业可提前准备相关内容，必要时，可引入富有经验的外部律师协助企业搭建算法合规体系，尽可能确保算法备案材料的齐备性。

4、备案期限

根据《算法推荐管理规定》第二十五条，当算法材料齐全的，网信部门应当在三十个工作日内予以备案。根据我们的项目实践，在深度合成算法备案过程中，可能存在需公司补正材料的情形，因此完成深度合成算法备案通常需2个月以上。

日前公布的《生成式人工智能服务管理暂行办法》亦对AIGC产品提出算法备案要求。作为人工智能领域的重要监管措施，算法备案已成为AIGC产品的入市门槛。鉴于算法备案所需时间较长，建议需履行算法备案的企业尽早开展算法备案工作。在具体的算法备案过程中，一方面，企业需对自身算法进行全面地梳理，由法务、IT协调配合准备相关备案材料，必要时可引入外部律师协助准备；另一方面，企业需充分了解相关法规要求和备案流程，必要时可与监管部门进行积极沟通和协调。完成算法备案后，应当及时在产品的显著位置公示备案编号，并以恰当的方式公示算法的基本原理、运行机制等内容。算法备案仅是算法治理的环节之一，完成算法备案并非一劳永逸，AIGC产业企业应当主动拥抱监管，积极履行《算法推荐管理规定》《深度合成管理规定》关于算法治理的其他合规义务。

/PART 003

结语

在AIGC产品上市的过程中，算法备案是保障其合规性的重要环节。在本篇中，我们深入探讨了算法备案的流程、内容以及注意事项。算法备案作为产品上市前的关键环节，不仅在法律法规层面上确保了AIGC产品的技术安全和输出内容安全，也是企业在境内人工智能赛道守法合规经营并保持长期发展的关键因素。随着算法备案机制的发展与完善，我们切实感受到算法备案的颗粒度与难度正在逐步增加，望负有算法备案义务的企业尽早启动相关工作。

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

跨越AIGC产品合规上市之路(二):
资质证照

ARTICLE BY 蔡鹏 肖莆羚令 王梦迪

对于任何一款AIGC产品来说，要在市场上合规运营，除了优秀的技术和功能，取得相应的资质证照也是至关重要的。资质证照不仅提供了法律和行业所需的合规性，而且也是企业信誉和品牌形象的象征。在一个充满不确定性和风险的数字环境中，拥有适当的资质证照是建立信任、保护权益、维护合法运营的重要条件。对于AIGC产品服务提供者来说，无论拟上市的产品是一个独立的网站、App还是小程序，都需要进行基础的资质申请。基础资质包括但不限于ICP备案、移动应用程序（App）备案、公安联网备案以及有关证书等。然而，这只是一款线上产品上市运营的基础条件之一。由于AIGC产品的特殊性质，往往涉及到更为复杂的业务场景和功能需求，相应需要取得如《网络文化经营许可证》《网络出版服务许可证》《信息网络传播视听节目许可证》或备案等资质证照。

本篇中我们将回归互联网产品的基本要求，结合我们的实践经验，重新审视AIGC产品相关资质证照的适用性。通过温故知新的分析，为相关服务提供方提供有价值的参考信息。

/PART 001

一般性资质

（一）互联网信息服务许可/备案（ICP许可/备案）

根据《互联网信息服务管理办法》的规定，国家对经营性互联网信息服务实行许可制度；对非经营性互联网信息服务实行备案制度。

- **经营性互联网信息服务，是指通过互联网向上网用户有偿提供信息或者网页制作等服务活动。实践中，常见的经营性互联网信息服务如网上广告、代制作网页、服务器硬盘空间出租、有偿提供特定信息内容、电子商务等。**

- **非经营性互联网信息服务，是指通过互联网向上网用户无偿提供具有公开性、共享性信息的服务活动，例如政府网站、在线论坛和社区、无偿线上图书馆或数据库等。**

针对AIGC产品而言，不论其产品形态是网站、App亦或是小程序，首先需要能够被公众通过域名或者IP地址访问到，才能够为用户提供进一步的服务。因此，获得ICP许可/备案是其上市的前提条件之一。

对于AIGC App的互联网信息服务资质问题，我们分别咨询了工信部以及北上广地区通管局，并整理答复如下：

监管机构	App的ICP备案	App的ICP许可证
工信部	工信部无专门针对App的ICP备案流程，App的ICP备案流程以各地通管局的要求为准。	App的ICP许可证申请事宜以各地通管局的要求为准。
北京通管局	**备案流程：**北京通管局当前的ICP备案流程中暂无专门针对App的ICP备案流程，但企业可选择申请网站备案，并在网站备案流程中选择App的ICP备案。 **备案客体：**根据App的访问途径不同，企业可对App对应的域名或者可访问App的IP地址进行备案。	北京通管局ICP许可证的咨询窗口为AI咨询，暂未通过咨询获取到AI类App是否需要取得ICP许可证的有效、准确信息。
上海通管局	**适用备案的判断标准：**ICP备案以是否使用境内的服务器为核心判断因素，而不论产品形式（App或者网站）。	AI类App目前暂无需取得ICP许可证，不论App中是否涉及提供付费业务。
广东通管局	**备案流程：**可按照通管局官网公布的流程申请App的ICP备案。通常服务器接入商（如三大运营商）会提供ICP备案的平台，由接入商对备案材料进行初审和复审，复审完成后，备案材料将会被提交至通管局，由通管局进行备案材料的终审。 **备案客体：**根据App的访问途径不同，企业可对App对应的域名或者可访问App的IP地址进行备案。	AI类App目前暂无需获取ICP许可证。

在上表基础上结合我们的实践经验，**当前AIGC产品以取得ICP备案为主流，不论其是否提供付费会员等付费服务。**

鉴于ICP备案管理的属地性及常态化，各地通管局官网基本均具备相应的线上备案指南及备案入口，建议AIGC产品服务提供者向属地通管局说明其产品具体服务形式，以获得通管局对于应当履行许可或备案程序的明确指示，并按照操作指南完成相应手续。

此外，ICP许可/备案服务经过二十余年的发展已经形成了一套完整、成熟且高度标准化的流程。在实际办理过程中，AIGC产品服务提供者可委托互联网接入服务商代为办理。

（二）移动应用程序（App）备案

根据《工业和信息化部关于开展移动互联网应用程序备案工作的通知》（以下简称"**《通知》**"）的相关要求，在境内从事互联网信息服务的App主办者，应当依照《中华人民共和国反电信网络诈骗法》《互联网信息服务管理办法》等规定履行备案手续，未履行备案手续的，不得从事App互联网信息服务。

结合我们的实践经验，针对AIGC产品而言，如产品形态为App、小程序，则App、小程序的主办者应当按照上述《通知》的相关要求，开展备案工作，并在完成备案后将备案号标注在App、小程序"设置"或"介绍"位置。

根据《通知》的相关要求，虽然备案的直接义务主体是App、小程序的主办者，但实际是由网络接入服务提供者、分发平台代为履行备案手续。当前，各大应用平台已发布了备案的具体流程及相关要求，结合《通知》的相关要求，可概括总结如下：

- 适用的产品形态：境内联网的App、小程序、快应用，不论是否向公众提供服务；
- 备案主体：App、小程序、快应用的主办者（此处的主办者等同于

信息服务提供者，根据《移动互联网应用程序信息服务管理规定》等，信息服务提供者包括**应用程序在法律上的所有者或者运营者**）；

- 备案时限：存量App、小程序、快应用应于2024年3月31日之前完成备案；[1]
- 备案渠道："国家互联网基础资源管理系统"（即ICP/IP地址/域名信息备案管理系统，网址为https://beian.miit.gov.cn/）；
- 备案流程：App、小程序、快应用主办者准备备案材料——通过网络接入服务商/应用平台相关系统入口提交备案申请，填写备案材料——网络接入服务商/应用平台初步核验（1-2个工作日）——工信部短信核验（初步核验完成提交通管局后，APP主办者根据短信提示，登陆https://beian.miit.gov.cn完成确认）——通管局审核（1-20个工作日）——结果下发（省级通管局审核通过后，通过短信、邮件下发备案号）——结果公示（工业和信息化部通过 https://beian.miit.gov.cn对备案结果进行公示）；

此外，需要提示AIGC产品服务提供者注意的是，移动应用程序（App）备案与上文提及的ICP备案虽均通过ICP/IP地址/域名信息备案管理系统（https://beian.miit.gov.cn）开展具体备案工作，但这两项备案是相互独立的，两者性质不同，需要提交的具体材料也存在差别，最终获得的备案号亦不相同。因此完成ICP备案后，AIGC产品服务提供者并非就此"高枕无忧"，仍需要按照移动应用程序（App）备案的相关要求在2024年3月31日前（还需关注不同应用平台的时限要求）完成备案工作。

（三）公安联网备案

根据《计算机信息网络国际联网安全保护管理办法》等相关规定，凡

1.部分开放平台提前了该时限要求，如小米开放平台在其《移动应用程序（App）备案指引》中明确2023年12月12号起，将限制未备案的库内应用发起版本更新；2024年1月4号起，将逐步开始清理在架未备案的存量应用。

是接入互联网的单位：包括互联网接入服务单位（ISP）、互联网数据中心（IDC）、互联网信息服务单位（ICP）和国际联网使用单位，均需到公安机关办理备案手续。根据行业实践，通常各网站在完成ICP备案后，或网站部署在非中国内地的服务器上但是为中国内地（大陆）提供服务时，需在网站开通之日起30日内登录"全国公安机关互联网站安全管理服务平台"提交公安联网备案申请。

针对AIGC产品而言，通常均需要通过互联网为用户提供信息服务，因此，AIGC产品网站的开办者或者App运营者均应当按照所在省、自治区、直辖市的属地公安机关的要求完成公安联网备案手续。

与ICP许可/备案类似，公安联网备案手续已经是所有网站、App的"基操"，备案材料、程序以及相关指引相对清晰明确。我们简单总结如下：

- 适用的产品形态：网站、App
- 备案主体：网站开办者、联网单位
- 备案时限：网络正式联通之日起30日内
- 备案渠道："全国公安机关互联网站安全管理服务平台"（https://www.beian.gov.cn/portal/index.do）
- 备案流程：注册——登录——开办主体信息审核（一般21个工作日内完成审核）——新办网站申请（一般21个工作日内完成审核）——非交互式网站初步审核后即完成备案/交互式网站面审或者实地检查——公安机关核发备案编号——将备案编号放置在网站首页下端

交互式服务，是指为用户提供向社会公众发布文字、图片、音视频服等信息的服务，包括但不限于论坛、社区、贴吧、文字、或者音视频聊天室、微博客、博客、即时通信、分享存储、第三方支付、移动应用商店等互联网信息服务。

- 流程时间：2-3个月

更多关于公安联网备案的具体操作流程，可参考"全国公安机关互联网站安全管理服务平台"中的《全国公安机关互联网站安全管理服务平台备案办事指南》。

（四）《计算机软件著作权登记证书》《App电子版权证书》或《软件著作权认证证书》

根据《计算机软件保护条例》，中国公民、法人或者其他组织对其所开发的软件享有著作权。软件著作权人可以向国务院著作权行政管理部门认定的软件登记机构办理登记。目前，国务院著作权行政管理部门认定的软件登记机构以及主流著作权认证有三类：

- 《计算机软件著作权登记证书》。该证书由中宣部下属机构中国版权保护中心登记颁发；
- 《App电子版权认证证书》。该证书由易版权平台——中国版权保护中心与北京版信通技术有限公司共建的移动App第三方证书签名与版权登记联合服务平台——作为认证机构，结合加密数字签名生成；
- 《软件著作权认证证书》由国家版权局主管的中国版权协会作审核认证，并使用区块链（"**中国版权链**"）进行唯一登记。

著作权本身并不因登记而取得，故上述三项证书均为非强制性的权利性证书，亦非AIGC产品上市的法定要件。但在当今的主流应用市场中，App软件权属成为应用市场的审核条件之一。对于AIGC产品服务提供者而言，在App产品进入应用市场环节的实践中，通常要求提供以上三种证书之一作为软件著作权归属证明。

因此，AIGC产品服务提供者基于满足上架条件的目的，应及时取得《计算机软件著作权登记证书》《App电子版权证书》或《软件著作权认

证证书》。此外，作为特别提示，上述三项证书的申请流程简单，费用不高，但企业需要注意申请材料的准备和齐全程度，避免因为不符合要求而需要进行额外的补正或重新提交申请。

/PART 002

基于业务场景的选择性资质

除上述必备要件外，AIGC产品的服务领域或业务场景可能导致服务提供者需要取得一些特殊执业证照才能够合法上市。结合我们的项目经验，以下几种资质证照为AIGC产品所常需：

（一）《网络文化经营许可证》

根据《互联网文化管理暂行规定》的规定，从事经营性互联网文化活动的企业应当申请《网络文化经营许可证》。经营性互联网文化活动是指以营利为目的，通过向上网用户收费或者以电子商务、广告、赞助等方式获取利益，提供互联网文化产品[2]及其服务的活动。

针对AIGC产品而言，基于其强大的产品功能及丰富的业务场景，产品将会根据用户的不同指令呈现不同的结果。如在虚拟数字人场景下，AIGC产品的功能中可能会涉及到虚拟数字人的表演、与用户之间的互动等。虚拟数字人的此类行为是否会构成网络表演进而需要申请《网络文化经营许可证》，是一个值得研究的问题。

2.根据《互联网文化管理暂行规定》第二条，“本规定所称互联网文化产品是指通过互联网生产、传播和流通的文化产品，主要包括：
（一）专门为互联网而生产的网络音乐娱乐、网络游戏、网络演出剧（节）目、网络表演、网络艺术品、网络动漫等互联网文化产品；
（二）将音乐娱乐、游戏、演出剧（节）目、表演、艺术品、动漫等文化产品以一定的技术手段制作、复制到互联网上传播的互联网文化产品。”

结合我们向相关监管部门的咨询，可以获悉的参考性判断标准为：应当根据AIGC产品的业务模式、网络文化产品（如网络表演、网络音乐）所占AIGC产品服务的比例、虚拟人物所提供服务的占比（零星或者大部分）来综合判断是否需要取得《网络文化经营许可证》。如AIGC产品的主要功能不是提供网络表演、网络音乐等网络文化产品，仅具备虚拟人物表演的小模块或者有零星的虚拟人物功能，暂不需要办理《网络文化经营许可证》。

基于《网络文化经营许可证》的属地管理属性，建议AIGC产品服务提供者在提供服务前，向省、自治区、直辖市人民政府文化行政部门进行咨询，明确是否需要申请相关许可证。

（二）《网络出版服务许可证》

根据《网络出版服务管理规定》的规定，从事网络出版服务，必须依法经过出版行政主管部门批准，取得《网络出版服务许可证》。网络出版服务是指通过信息网络向公众提供网络出版物。网络出版物，是指通过信息网络向公众提供的，具有编辑、制作、加工等出版特征的数字化作品，范围主要包括：（1）文学、艺术、科学等领域内具有知识性、思想性的文字、图片、地图、游戏、动漫、音视频读物等原创数字化作品；（2）与已出版的图书、报纸、期刊、音像制品、电子出版物等内容相一致的数字化作品；（3）将上述作品通过选择、编排、汇集等方式形成的网络文献数据库等数字化作品；（4）国家新闻出版广电总局认定的其他类型的数字化作品。

对于AIGC产品而言，AIGC的可版权性是当下讨论的热点问题，AI生成的内容是否享有版权以及版权归属于哪一主体，也相应影响着AIGC产品服务提供者是否会构成网络出版服务提供者。结合当前国家新闻出版署关于"设立网络出版服务单位审批"的公示，网络出版服务许可主要还是聚

焦在针对图书、音像、电子、报纸、期刊出版单位从事网络出版服务方面，AIGC产品的功能场景中直接涉及到网络出版服务的场景较少。

基于网络出版服务许可管理的属地性，建议AIGC产品服务提供者在提供服务前，向所在地省级出版行政主管部门进行咨询，明确是否需要申请相关许可证。

（三）《信息网络传播视听节目许可证》

根据《互联网视听节目服务管理规定》的规定，从事互联网视听节目服务，应当依照本规定取得广播电影电视主管部门颁发的《信息网络传播视听节目许可证》（仅国有独资或国有控股单位才具备申请许可证的资质条件）或履行备案手续。（地（市）级以上广播电台、电视台、中央新闻单位提供互联网视听节目转播类服务的，无需申请许可证但需要履行备案手续）。

对于AIGC产品而言，目前产品功能基本不涉及提供如新闻、广播、转播等服务，因此需要进行信息网络传播视听节目备案手续的可能性较小。鉴于AIGC产品发展的快速性及功能场景的多样性，后续如果在垂直领域中予以应用，则不排除需要备案的可能性。AIGC产品服务提供者可结合原国家新闻出版广电总局（现国家广播电视总局）发布的《互联网视听节目服务业务分类目录（试行）》[3]判断相关产品功能是否构成互联网视听节目服务业务。

3.《互联网视听节目服务业务分类目录（试行）》：http://gbdsj.gd.gov.cn/attachment/0/399/399935/3072082.pdf

/PART 003

结语

综上所述，AIGC产品在上市运营前需要取得多种资质证照。这些资质证照的取得既是保护企业合法运营的基础条件，也是保障企业自身权益的重要手段。对于AIGC产品服务提供者来说，积极申领相关资质证照，取得上架“通行证”，是其在竞争激烈的市场中站稳脚跟，获得用户的信任和支持的必备要件。

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

谨防“假作真时真亦假”

——生成式人工智能的真实性问题及治理

ARTICLE BY 王红燕

本文从生成式人工智能生成内容的真实性问题出发，深度解析了我国法律对生成式人工智能技术真实性的监管规则，并对AIGC技术相关企业的合规治理提出了建议，以供读者参考。

/PART 001

生成式人工智能生成内容的真实性问题

生成式人工智能（AI Generated Content，AIGC），是指基于算法、模型、规则生成文本、图片、声音、视频、代码等内容的技术。[1]2022年12月，生成型预训练变换模型（Chat Gegerative Pre-trained Transformer，ChatGPT）火爆出圈，引发了对AIGC技术的新一轮关注。与分析已有数据的预测性人工智能不同,生成式人工智能可以通过学习海量数据来生成新的数据、语音、图像、视频和文本等内容。由于AIGC技术本身不具备判断力，随着AIGC技术的应用越来越广泛，其可能生成的虚假信息所带来的弊端也日益严重。不少用户在使用ChatGPT时已经意识到，ChatGPT的回答可能存在错误，甚至可能无中生有地臆造事实，臆造结论，臆造引用来源，虚构论文、虚构新闻等。面对用户的提问，ChatGPT会给出看似逻辑自恰的错误答案。在法律问题上，ChatGPT可能会虚构不存在的法律条款来回答问题。

OpenAI在GPT-4技术报告中指出，GPT-4和早期的GPT模型（包括大家熟知的ChatGPT）生成的内容并不完全可靠，可能存在"Hallucinations"（臆造），即"产生与某些来源无关的荒谬或不真实的内容"。[2]随着GPT模型越完善越智能，用户将更难区分其生成内容是真实的还是虚构的，并且，GPT模型生成的虚假数据极有可能被再次"喂养"给机器学习模型，致使虚假信息进一步泛滥，用户被误导的可能性进一步增大，而获得真实信息的难度增加。知名问答网站Stack Overflow就发布临时政策禁止ChatGPT在网站中的使用，以应对ChatGPT生成内容的泛滥之势，因为

1.国家互联网信息办公室，《生成式人工智能服务管理暂行办法》
2.OpenAI（2023），《GPT-4 Technical Report》

这些内容漏洞百出，质量低下，会给来网站寻求帮助的用户造成严重困扰和不便，严重影响平台内容质量。

因此，AIGC技术生成内容的真实性问题不仅是一个需要解决的技术问题，也需要通过法律制度加以监管和治理。

/PART 002
我国法律对生成式人工智能技术真实性的监管

早在2019年11月18日，国家互联网信息办公室,文化和旅游部,国家广播电视总局三部委发布的《网络音视频信息服务管理规定》就对基于深度学习、虚拟现实等的新技术新应用提出了监管要求，包括不得用于制作、发布、传播虚假新闻信息，制作、发布、传播非真实音视频信息应当以显著方式标识；发现网络音视频信息服务使用者利用基于深度学习、虚拟现实等的虚假图像、音视频生成技术制作、发布、传播谣言的，应当及时采取相应的辟谣措施，并将相关信息报网信、文化和旅游、广播电视等部门备案；具有媒体属性或社会动员功能的应当开展安全评估。

2021年12月31日，国家互联网信息办公室,工业和信息化部,公安部,国家市场监督管理总局四部委发布了《互联网信息服务算法推荐管理规定》（以下简称**《算法推荐管理规定》**），将应用算法推荐技术界定为“利用生成合成类、个性化推送类、排序精选类、检索过滤类、调度决策类等算法技术向用户提供信息”，要求算法推荐服务提供者不得生成合成虚假新闻信息，不得传播非国家规定范围内的单位发布的新闻信息。具有舆论属性或者社会动员能力的算法推荐服务提供者应完成算法备案，开展安全评估。不得利用算法虚假注册账号、非法交易账号、操纵用户账号或者虚假点赞、评论、转发，不得利用算法屏蔽信息、过度推荐、操纵榜单或者检索结果排序、控制热搜或者精选等干预信息呈现，实施影响网络舆

论或者规避监督管理行为。

2022年11月25日，国家网信办等三部委发布的《互联网信息服务深度合成管理规定》（以下简称**《深度合成管理规定》**）在《网络音视频信息服务管理规定》和《算法推荐管理规定》的基础上进一步提出对深度合成技术的细化监管要求，规定"不得利用深度合成服务制作、复制、发布、传播虚假新闻信息"。2023年2月，一条关于杭州取消限行的消息在网络上广泛传播，此事惊动警方介入调查，最终发现这是一则ChatGPT生成的假新闻，是有人在业主群展示ChatGPT功能时闹出的乌龙事件。

从"杭州取消限行乌龙事件"可以看到，虽然《网络音视频信息服务管理规定》《算法推荐管理规定》《深度合成管理规定》或多或少可以对制作、发布、传播虚假新闻信息等行为进行监管，但无法涵盖使用例如ChatGPT等AIGC技术生成除虚假新闻之外虚假内容的行为，考虑到AIGC技术引发的虚假信息风险，在前述规定下恐怕暂时难以得到恰当的解决。

2023年7月10日，国家网信办等部门共同发布了《生成式人工智能服务管理暂行办法》（以下简称**《AIGC服务管理办法》**），在《网络音视频信息服务管理规定》《算法推荐管理规定》《深度合成管理规定》的基础上，制定了多条规定对AIGC生成内容的真实性进行特别强调，包括《AIGC服务管理办法》第四条规定"提供和使用生成式人工智能服务，应当遵守法律、行政法规，尊重社会公德和伦理道德，遵守以下规定：

（一）坚持社会主义核心价值观，不得生成煽动颠覆国家政权、推翻社会主义制度，危害国家安全和利益、损害国家形象，煽动分裂国家、破坏国家统一和社会稳定，宣扬恐怖主义、极端主义，宣扬民族仇恨、民族歧视，暴力、淫秽色情，以及虚假有害信息等法律、行政法规禁止的内容……（五）基于服务类型特点，采取有效措施，提升生成式人工智能服务的透明度，提高生成内容的准确性和可靠性。《AIGC服务管理办法》第七条规定"生成式人工智能服务提供者（以下称提供者）应当依法开展预

训练、优化训练等训练数据处理活动，遵守以下规定：……（四）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"等。

从《网络音视频信息服务管理规定》《算法推荐管理规定》《深度合成管理规定》的相关规定中，笔者倾向于认为，针对基于深度学习等技术生成或合成的内容，我国较倾向于管控具有舆论导向性、媒体属性或社会动员功能的生成内容，尤其是虚假新闻、颠覆国家政权、危害政治安全和社会稳定等方面的信息。最新发布的《AIGC服务管理办法》则更广泛要求生成式人工智能生成的内容不得含有虚假信息，提供生成式人工智能产品或服务应当采取措施防止生成虚假信息。显然，《AIGC服务管理办法》中的虚假信息涵盖了虚假新闻以及其他臆造的、不真实的信息。

实际上，从技术角度看，正是因为生成式人工智能与预测性人工智能不同,能够生成新的数据、语音、图像、视频和文本等内容，许多生成式人工智能的生成内容是发散的、带有创新性的，甚至可能突破固有思维，从而使得其能够被人称之为"革命性技术"，这也意味着要求生成式人工智能的生成内容需确保真实性与生成式人工智能的技术原理存在一定的冲突。法律与技术的关系并非简单的治理与被治理的关系，《AIGC服务管理办法》发布的目的也在于促进生成式人工智能健康发展和规范应用，因此，法律监管的力度需要考虑技术的更新迭代速度，前瞻布局，为未来新技术的监管留有空间，同时，生成式人工智能服务的提供者需要在法律法规范围内推动技术的发展。

/PART 003
对AIGC技术相关企业的合规治理建议

在前述法律法规中，针对不真实的信息有些采用了"谣言"、"网络谣

言"的表述，有些采用"虚假信息"这样的表述，虽然这些表述目前在任何法律法规的条文中都尚没有非常精确的定义。从法理上来说，谣言或虚假信息的本质是虚假的、缺乏事实依据的信息。制造、传播谣言或虚假信息是否构成违法犯罪以及应当施以什么样的处罚，与该行为所造成负面影响的程度有关。例如，《刑法》第二百四十六条规定了诽谤罪。而当同一诽谤信息实际被点击、浏览次数达到5000次以上，或者被转发次数达到500次以上的，就可被认定为"情节严重"。虽然单条、少数或传播范围有限的谣言或虚假信息的影响较小，但考虑到网络传播的无序性、快速性、不受地域限制等特点，容易使点击次数、浏览次数达到入刑标准。

目前《AIGC服务管理办法》第九条明确规定，"提供者应当依法承担网络信息内容生产者责任，履行网络信息安全义务。涉及个人信息的，依法承担个人信息处理者责任，履行个人信息保护义务。提供者应当与注册其服务的生成式人工智能服务使用者（以下称使用者）签订服务协议，明确双方权利义务。"因此，针对AIGC生成内容的真实性问题，为防范法律风险，提前规避责任，笔者建议结合国内相关企业应当结合所提供的技术服务，在现行法律法规的框架下进行有针对性的合规治理。

涉及网络音视频服务、算法推荐服务和深度合成服务的生成式人工智能提供者均应当至少做好以下共性的合规工作：

（1）建立信息发布审核制度和技术措施，维护数据的完整性、安全性和可用性；

（2）对用户进行真实身份信息认证，对不提供真实身份信息的用户，不为其提供信息发布服务；

（3）对生成合成的内容进行显著标识；

（4）建立辟谣机制。发现利用制作、复制、发布、传播虚假信息的，应当及时采取辟谣措施，保存有关记录，并向网信部门和有关主管部门报告；

（5）设置便捷的投诉举报入口，公布投诉、举报方式等信息，及时受

理并处理公众投诉举报；

（6）提供具有舆论属性或者社会动员功能的技术服务的，按照国家规定开展安全评估；

（7）提供服务前，履行备案和变更、注销备案手续，完成备案的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；

（8）发现、知悉生成的内容不符合要求时，应当及时采取处置措施；

（9）对于运行中发现、用户举报的不符合要求的生成内容，除采取内容过滤等措施外，应在3个月内通过模型优化训练等方式防止再次生成。

针对不同的技术服务，网络音视频服务提供者，还应当注意部署应用对违法违规音视频和非真实音视频的鉴别技术。

算法推荐服务提供者还应当注意：（1）建立健全用于识别违法和不良信息的特征库，完善入库标准、规则和程序，记录并留存相关网络日志；（2）提供互联网新闻信息服务的，应当依法取得互联网新闻信息服务许可；（3）依法开展涉电信网络诈骗信息的监测、识别和处置。

深度合成服务提供者，针对提供（1）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；（2）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；（3）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；（4）沉浸式拟真场景等生成或者编辑服务等可能导致公众混淆或者误认的几类特殊场景的，要求深度合成服务提供者应当在生成或者编辑的信息内容的合理位置、区域进行显著标识，向公众提示深度合成情况。而除此几项特殊场景的深度合成服务提供者，则应当提供显著标识功能，并提示深度合成服务使用者可以进行显著标识。

由此，通过用户身份认证，可以删减部分带有不良目的的用户，从源头减少虚假信息的产生；通过信息发布的审核机制，可以通过技术手段或

人工方式剔除虚假信息，保证生成数据的安全性和可用性；通过显性标识，可以实现虚假信息的溯源，提供追责线索；通过辟谣机制、投诉通道、处置措施，为因虚假信息所引发的问题提供了解决和纠正途径。

目前，最新发布的《AIGC服务管理办法》对AIGC技术提出了一些较为严格的合规要求，AIGC相关技术企业均已开始着手开展合规工作，加强自律自治，加强行业制度规范的制定，共同打造良好的AIGC产业生态。以某短视频社交平台为例，其于2023年5月9日提出十一条平台规范暨行业倡议：一方面，该社交平台对参与平台生态的创作者、发布者、用户、商家、广告主等主体提出要求，包括（1）发布者应对人工智能生成内容进行显著标识，帮助其他用户区分虚拟与现实，特别是易混淆场景。（2）发布者需对人工智能生成内容产生的相应后果负责，无论内容是如何生成的。（3）禁止利用生成式人工智能技术创作、发布违背科学常识、弄虚作假、造谣传谣的内容。一经发现，平台将严格处罚；另一方面，该社交平台通过平台内部的技术力量协助对AIGC生成内容涉及的问题进行治理，以保护用户权益，包括（1）平台将提供统一的人工智能生成内容标识能力，帮助创作者打标，方便用户区分。（2）平台将提供用户反馈渠道，方便用户反馈违规生成内容等条款。[3]AIGC相关技术企业在遵照《AIGC服务管理办法》进行合规制度建设的同时，仍需紧密关注法律政策动态，以期在合规合理的范围内更好地应用生成式人工智能技术。

（陈茜对本文亦有贡献）

3.《某短视频社交平台关于人工智能生成内容的平台规范暨行业倡议》，来源于https://www.douyin.com/rule/billboard?id=1242800000049，最后访问时间2023年5月18日。

王红燕
合伙人
知识产权部
杭州办公室
+86 571 5662 3968
gracewang@zhonglun.com

A

CHAPTER

05

侵权责任

C

G

机器学习作品的类型化及其著作权责任

ARTICLE BY 王红燕

随着人工智能的智能化程度越来越高，传统视野下的著作权制度遭到了前所未有的挑战。以深度学习算法为核心的人工智能系统实现了引人瞩目的成就，比如腾讯写作机器人Dreamwriter撰写的财经报道，专业人员认为与媒体记者日常的消息稿无异[1]，微软人工智能产品小冰于2017年独立创作并出版了诗集《阳光失了玻璃窗》[2]，微软小冰又于2019年在中央美术学院美术馆展出首个个展《或然世界》[3]。目前学界和实务界关于人工智能输出端的生成物是否是作品以及归属于谁的话题热度不减，而对人工智能训练过程中输入端的机器学习作品是否涉嫌侵犯著作权的问题则讨论不多。人工智能的前期训练是人工智能"创作"的必要条件，而目前主流的训练算法以深度学习为核心，并且需要海量的训练数据作为人工智能学习和成长的"养料"，人工智能训练团队对海量数据的获取以及输入，势必存在侵犯著作权的风险。因此，对于以输入海量作品作为训练数据的机器学习是否有侵权之嫌、是否可以适用合理使用制度提出侵权抗辩以及中国当前著作权体系是否能够对这种新的行为进行认定是本文要探讨的内容。

1.人民网：腾讯开发新闻写作机器人，记者们是否已哭晕？http://it.people.com.cn/n/2015/0911/c1009-27570800.html，2022-4-3

2.中国青年网：人工智能微软小冰出首部诗集http://news.youth.cn/jsxw/201706/t20170601_9938618.htm，2022-4-3

3.澎湃新闻：微软AI小冰在中央美术学院举办首个画展：名叫《或然世界》https://baijiahao.baidu.com/s?id=1638953179316720516&wfr=spider&for=pc，2022-4-3

/PART 001

机器学习的概念和类型划分

机器学习的主要研究对象是人工智能，是对能通过经验自动改进的计算机算法的研究。深度学习，是指采用深度模型进行机器学习的学习方法，它学习的是样本数据的内在规律和表示层次。[4]

1、机器学习怎么"喂养"数据？

以深度学习为核心算法、海量训练数据为学习材料的机器学习是人工智能的智能化程度不断提升的基础，而包括著作权作品在内的数据"喂养"，会面临侵犯著作权的风险。以微软小冰创作诗集为例，来简要说明一下机器学习的过程：

微软小冰训练团队将1920年代起到现在的519位中国现代诗人的几万首诗歌，运用图像识别等技术数字化为计算机可读的语言，作为微软小冰训练的语料库，输入到微软小冰的诗歌生成模块中进行训练，训练人员可以设置相应的训练次数，相应训练结束后，训练人员通过诱发模块给出创作诱发信号，按照各项指标评价微软小冰训练相应次数以后创作出的诗歌，将评价信息作为反馈来完善诗歌生成模块，当训练人员认为小冰经过一定次数训练以后创作出的诗歌具有一定美感时才会停止训练。[5]我们发现，从微软小冰零基础学习写诗到创作出具有一定美感的诗歌这一过程中，可能涉及享有著作权作品的输入、对著作权作品的改编或汇编性输出，因此，这一过程中存在侵犯作品的复制权、演绎权等法律风险。

4.百度百科：机器学习
https://baike.baidu.com/item/%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0/217599?fr=aladdin，2022-4-3
5.新智元：微软小冰被训练成诗人，人类或找到AI创造的通用方法
https://cloud.tencent.com/developer/article/1075746，2022-4-3

2、机器学习怎么分类？

复制权是一种依附性权利，控制复制行为的目的在于控制后续的传播和使用行为，所以，训练数据"喂养"过程中以传播效果为导向的复制很可能落入复制权的规制范畴，基于此，可以表达性内容的输出与否作为标准，将使用行为划分为表达性使用和非表达性使用。[6]

相应地，以是否有表达性内容输出为标准，将机器学习分为表达型机器学习和非表达型机器学习两类，并以机器学习的作品是否来源于特定作者为标准，将表达型机器学习进一步地划分为普通的表达型机器学习和特殊的表达型机器学习。[7]

(1)非表达型机器学习，指没有表达性内容输出的机器学习。此类典型的人工智能系统如人脸识别系统，人脸识别系统以训练人员输入的人脸照片作为训练材料，经过深度学习完成人脸照片像素点阵化——提取面部特征值——构建对应的特征值数字矩阵的识别算法训练，完成训练的人脸识别系统再基于识别算法对现实场景的识别需求作出回应。

(2)普通的表达型机器学习，指有表达性内容输出的机器学习，且用于算法学习的材料不局限于某一类特定作品，而来源于不特定的作品。比如，微软小冰将中国近现代五百多位诗人的诗歌作为语料库来训练诗歌生成模块，诗歌生成模块运用双向语言模型根据诱发源提取的多个关键词扩展成诗句，经过整诗的流畅性与连贯性检查后，输出创作的现代诗歌。[8]

(3)特殊的表达型机器学习，指有表达性内容输出的机器学习，且用于算法学习的材料来源于特定的作者。比如，微软将勃朗特将近十七万幅作品的片段作为人工智能系统的训练材料，提取作品片段的绘画细节和绘

6.参见高佳佳：《类型化视角下机器学习的合理使用分析》，《电子知识产权》2021年第5期
7.参见李安：《机器学习作品的著作权法分析——非作品性使用、合理使用与侵权使用》，《电子知识产权》2020年第6期
8.Wen-Feng Cheng, Chao-Chung Wu, et al. Image Inspired Poetry Generation in Xiaolce. arXiv preprint arXiv:1808.03090. 2018

画风格训练人工智能系统创作模块，最终人工智能系统创作出与勃朗特风格近似但是不相同的作品。[9]

/PART 002

著作权合理使用之三步检验法与转换性使用

著作权法赋予著作权人的专有权利并不是一种绝对控制其所创作的作品的权利，因此，在鼓励创作者创作和促进公众获得作品这两种利益的平衡下，著作权法对专有权利在一定程度上加以限制，作出合理的例外规定。这种合理的例外规定在国际条约上体现为：

《伯尔尼公约》第9条第2款规定：成员国法律有权允许在某些特殊情况下（不经作者许可）复制作品，只要这种复制不致损害作品的正常使用，也不致无故侵害作者的合法利益。同样，在TRIPs协定和《世界知识产权组织版权条约》的规定中，条约成员国也可以对作品的专有权作出合理的限制。

（一）比较法视角看著作权合理使用的立法规制

作为国际条约的成员国，成员国国内立法中对于著作权加以限制的例外规定必须以国际条约为前提。各国立法在著作权的限制和例外上的名称和体例有很大区别，主要分为两类[10]：

1、以美国为代表的一般条款型

美国国内立法《版权法》第107条并没有以列举的形式规定"合理使用"对应的各种情形，只给出法官认定行为是否构成"合理使用"的四个考

9.许建，朱韶斌：谁是《下一个伦勃朗》的作者？
https://www.zhichanli.com/p/749200061
10.王迁：《知识产权法教程》（第7版），第284页

量因素：（1）使用的目的和性质，即使用是出于商业目的还是教育目的；（2）被使用作品的性质；（3）被使用部分的数量和重要性；（4）对作品潜在市场或价值的影响。

2、以欧洲大陆法系国家为代表的列举穷尽型

欧洲大陆法系国家的著作权立法通常不以"一般条款"的形式赋予法官运用自由裁量权认定行为是否构成合理使用，而是制定"权利的例外和限制条款"，对是否构成合理使用的情形作出封闭性的全面列举，对不属于列举情形的行为没有适用合理使用的可能性。

我国《著作权法》的合理使用制度沿用了大陆法系国家的立法模式，但是所有区别。2010年的《著作权法》列举了合理使用的12种情形，随着这种封闭性的列举形式逐渐不适应司法实践的需求，2020年修订的《著作权法》增加了"法律、行政法规规定的其他情形"的兜底情形，在穷尽12种列举的法定情形下，给法官留下了合理使用认定上自由裁量权的开口。

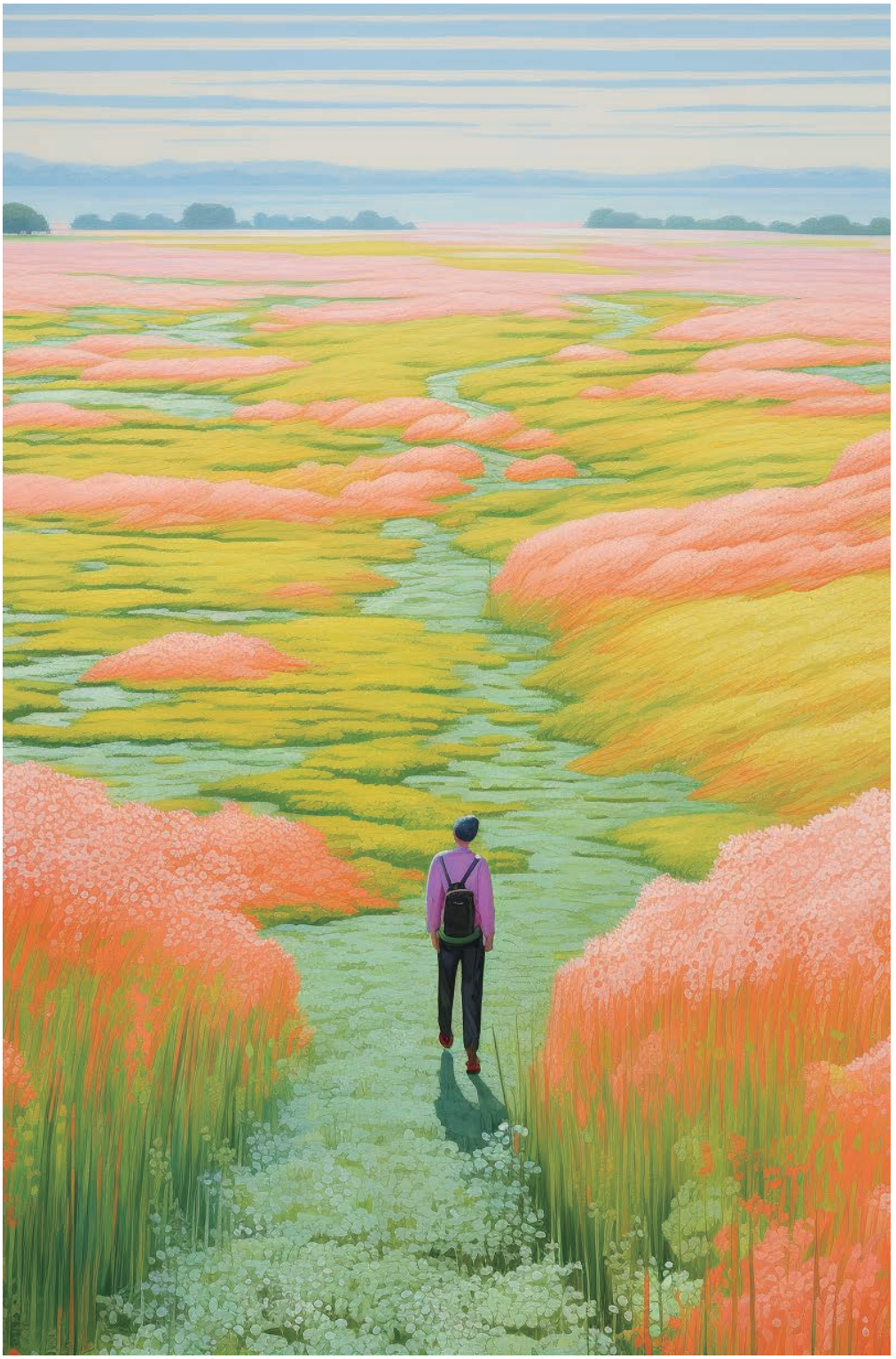
（二）著作权合理使用的认定模式

各国著作权合理使用的立法虽然存在体例和名称上的不同，但是，合理使用条款背后的立法精神确是统一的，无非是各国按照本国国情和司法实践对条款的各项要件的解释偏向有所区别。对于包括机器学习作品行为的合理使用认定的法律适用中，法官主要有以下两种论证模式：

1、三步检验法

(1)特殊且特定情形

该要件对应到我国《著作权法》的合理使用条款，法官在认定未经著作权人同意，使用作品的行为是否构成合理使用的免责情形时，仅限定于列举的十二种法定情形及增加的"其他情形"。有学者认为，《著作权法》合理使用条款在明文列举的合理使用情形之外增加了一项 "其他情形"作



为开放性司法解释的入口，实际上违背了"三步检验法"中特定且特殊情形的要件，因为"特定"意指著作权例外类型应由法律明确界定，而不能仅提供模糊标准。[11]

在我国法律体系中主要有三部著作权领域相关的规范性文件涉及"法律、行政法规规定的其他情形"，即《计算机软件保护条例》《信息网络传播权保护条例》和《著作权法实施条例》等，而随着技术革新以及法律的滞后性，新类型化的行为，包括机器学习作品的行为，是否构成合理使用将会面临没有法律上的认定依据的境地。

司法实践中，为应对法律失位、无法依法裁判的情况，最高人民法院2011年发布的《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："妥当运用著作权的限制和例外规定，正确判定被诉侵权行为的合法性……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……"，这份由最高院制定发布的司法文件，突破了《著作权法》的法定列举情形的限制，实际上是引入了美国《著作权法》对于合理使用行为考量四要素，在司法实践中起到了很好的效果，如谷歌图书数字化及片段式使用案[12]。

（2）不影响作品的正常使用

"不影响正常使用"要件采取了"禁止竞争性经济利益标准"，即要求合理使用行为不得与法定权利行使所获经济利益相冲突，所有法定权利所生

11.参见熊琦：《著作权转换性使用的本土法释义》，《法学家》2019年第2期
12.参见北京市第一中级人民法院(2011)一中民初字第1321号民事判决

成的收益应归属于著作权人所有。[13]该要件保护的是著作权人在权利行使过程中产生的经济利益，但是，著作权与经济利益的因果性，以及何种经济价值所包含的经济利益是《著作权法》在排除专有权的绝对控制以外所应当赋予著作权人的法益，都是司法实践中认定这一要件的难题。

另外，正常使用过程中产生的经济利益应当包括现有的和预期的利益。针对直接使用著作权作品的情形，如涉及侵犯复制权、信息网络传播权的行为，司法实践中常常以新作品是否产生了对原作品市场的替代性效应来认定行为的合法性，而针对涉及原作品的演绎权而使用作品的行为，虽然演绎作品改变原作品的表达方式，并且可能在与原作品不同的市场产生经济利益，但是，这类经济利益通常可以视为原作品正常使用过程中产生的预期的经济利益。

(3)不得不合理损害合法利益

"正常使用"从文义上讲指通常行使的权利，覆盖了过大范围的著作权市场，导致新技术发展迅猛的今天没有了新市场和新价值的空间，并且从理论上来讲，任何转换性使用行为都必然会对作品的市场价值造成影响，所以，解释"不得不合理损害合法利益"要件的重点，在于对"不合理"界限的确定。[14]不合理损害和正常使用考量的因素都在于经济利益，所以，不合理损害的利益同样延伸至预期市场，而对不合理损害的认定也在于是否对原作品市场产生替代性，以及对原作品的使用行为是否构成转换性使用等考量因素。

2、转换性使用

美国《版权法》第107条规定了合理使用，条文如下：

出于例如批评、评论、新闻报道、教学（包括供教室教学的多件复

13.参见熊琦：《著作权转换性使用的本土法释义》，《法学家》2019年第2期
14.参见熊琦：《著作权转换性使用的本土法释义》，《法学家》2019年第2期

制）、学术或研究等目的……对于受版权保护的作品的合理使用，不属于版权侵权。在判断对于作品的使用在某种情况下是否构成合理使用时需要考虑以下因素：

（1）使用的目的和性质，包括是否出于商业目的或非营利的教育目的；

（2）受到版权法保护的作品的性质；

（3）被使用部分的数量和重要程度对于被使用的作品的整体的情况；

（4）这种使用对于被使用作品的潜在市场或者作品的价值的影响。

如果对于作品的使用经过上述因素的判断可以认定为合理使用，则作品还未发表的事实本身不会影响合理使用的成立。

对于是否构成合理使用四个要素的认定，美国法院判例存在不同观点，本质是基于原作品产生新价值和保护原作品合法利益的矛盾。美国联邦最高法院在"索尼案"中明确提出，在没有证据证明的前提下，对有版权保护的作品的二次商业性使用被推定为不合理使用。[15]美国联邦最高法院在"坎贝尔案"中强调，商业性使用对于认定合理使用与否并不具有决定性，只是作为衡量合理使用的第一个因素。当二次作品越具有"转换性"，其他阻碍合理使用认定的因素（比如商业性使用）的重要性就会越小。[16]美国第二巡回上诉法院在"谷歌图书案"中认为，如果复制原作品之后创作的新作品具有高度转换性、创造性，并且新作品产生的市场不构成对原作品受保护的市场实质替代，谷歌的商业性使用行为以及使用比例占原作品百分比过高都不能作为否定合理使用的正当理由。[17]

所谓"转换性使用"，是指对原作品的使用并非为了单纯地再现原作品

15.Sony Corp. v. Universal City Studios，464 U.S.417，1984
16.Campbell v. Acuff-Rose Music, Inc.,510 U.S.569，1994
17.Authors Guild v. Google Inc., 804F. 3d 202（2nd,2015）

本身的文学、艺术价值或者实现其内在功能或目的，而是通过增加新的美学内容、新的视角、新的理念或通过其他方式，使原作品在被使用过程中具有新的价值、功能或性质，从而改变了其原先的功能或目的。[18]

具体有两种典型的转换性使用行为：

（1）转换内容的使用行为。其中涉及的转换主要集中于以批注、评论或再创作的方式对原作品加以改动。[19]我国首例转换性使用判决中，法官认为，“黑猫警长”等美术作品被引用在电影海报中具有了新的价值、意义和功能，其原有的艺术价值功能发生了转换，而且转换性程度较高，不会产生替代性使用，亦不会影响权利人的正常使用。[20]

（2）转换目的的使用行为。不改变作品表达，仅改变作品使用目的的行为，也被称为功能性转换。[21]在美国“谷歌图书”案中，美国法院对谷歌公司的复制行为以及片段引用行为均作出合理使用的认定，但我国法院在谷歌图书数字化及片段式使用案中就相似的案件事实，一审法院仅认为片段引用行为属于转换性使用，而复制行为被认定为侵权[22]，二审法院维持原判，但指正了一审法院对谷歌公司的复制行为是否构成侵权的多项说理，认为“如果是专门为了后续的合理使用行为而未经许可复制他人作品，应当认定为合理使用行为的一个部分,同样构成合理使用”，二审法院基于谷歌公司对合理使用的抗辩提供证据不足驳回了上诉，并未在实体法上对谷歌公司的复制行为是否构成合理使用作出评价[23]。

18.王迁：《知识产权法教程》（第7版），第330页
19.参见熊琦：《著作权转换性使用的本土法释义》，《法学家》2019年第2期
20.参见上海知识产权法院(2015)沪知民终字第730号民事判决
21.参见熊琦：《著作权转换性使用的本土法释义》，《法学家》2019年第2期
22.参见北京市第一中级人民法院(2011)一中民初字第1321号民事判决
23.参见北京市高级人民法院（2013）高民终字第1221号民事判决

/PART 003

机器学习作品著作权责任

1、非表达型机器学习

非表达型机器学习仅有著作权作品的输入，以人脸识别系统的学习训练来看，对版权图片提取特征值并构建特征矩阵，并不涉及对作品创造性内容的提取，因此，不是表达性内容的输入，且后续训练完成的人工智能系统的应用场景也不会有表达性作品的输出，这种机器学习训练中使用作品的行为不属于著作权法意义上的作品使用，不构成著作权侵权。

2、普通的表达型机器学习

该类机器学习作品强调不特定作品的输入，比如，在微软小冰诗歌生成模块的训练中，输入五百多位诗人的诗歌作为语料库来训练小冰根据关键词扩展成诗句的能力，训练过程主要是提取原作品的高频词组表达，创作出的新作品也是根据概率分布和高频搭配等方法的词语扩展，很难形成对原作品的替代性使用，同样也不会损害原作品的市场利益，因此，可以适用合理使用进行侵权抗辩。

3、特殊的表达型机器学习

在微软的勃朗特绘画创作人工智能系统中，仅输入勃朗特三百多幅作品，并训练人工智能系统识别和分析勃朗特的绘画风格和绘画细节，包含了对表达性作品的独创性表达的提取，同时，输出的新作品具有强烈的勃朗特风格，会对原作品的市场产生替代性效应，不属于作品的正常使用，因此，该类机器学习作品不能用合理使用进行抗辩。

/PART 004

结语

人工智能技术的发展不仅在于算法、算力、数据三要素的提升，也与著作权法对人工智能技术各个环节的规制与引导息息相关。本文将人工智能机器学习作品行为类型化，并对合理使用条款法律适用进行解析，在人工智能时代司法实践中有一定意义。

（陈茜、宋凯辉对本文亦有贡献）

王红燕
合伙人
知识产权部
杭州办公室
+86 571 5662 3968
gracewang@zhonglun.com

人工智能时代涉数据、算法的新型不正当竞争行为及法律规制

ARTICLE BY 王红燕

在人工智能时代，经营者可能利用先进算法与大数据资源实施复杂、隐秘的不正当竞争行为，对其他经营者合法提供的网络产品或服务产生妨碍干扰甚至是颠覆性影响。相关研究表明，到2020年，零售业85%的顾客服务互动将由某种形式的人工智能技术驱动或受其影响。来自广告公司JWalter Thompson的一份报告表明，70%的所谓千禧一代欣赏通过人工智能技术展示其产品的品牌，38%的消费者在使用人工智能的情况下获得了比不使用人工智能更好的购物引导。[1]然而在与人工智能相关的法律问题研究方面，鲜少有围绕人工智能技术与反不正当竞争展开讨论。

1. Lee Curtis & Rachel Platts. AI Is Coming and It Will Change Trade Mark Law [J]. Managing Intellectual Property, 2017(1):10.

/PART 001

人工智能时代涉数据、算法的新型不正当竞争行为

人工智能的三大基础要素为数据、算法、算力。虽然数据、算法本身是中立的技术手段，但经营者具有"经济人"和"理性人"双重属性，可能会滥用数据、算法，突破法律和商业伦理的边界，实施流量劫持、妨碍干扰、恶意不兼容等新型不正当竞争行为，最终损害其他市场参与方的合法权益。

不正当竞争的成立可以是损害经营者法定的有名权益，如商标、商业秘密等，也可以是无名权益，只要其可以作为一种竞争优势给经营者带来营业收入或潜在交易机会。涉数据、算法的新型不正当竞争行为是指经营者利用数据、算法实施的违反法律与商业道德的竞争行为。在被称为中国大数据产品不正当竞争第一案的T公司诉M公司案[2]中，T公司对于"生意参谋"数据产品享有竞争性财产权益。法院认为，M公司以营利为目的，将"生意参谋"数据产品直接作为自己获取商业利益的工具，提供同质化的网络服务，从而获取商业利益与竞争优势的行为，明显有悖公认的商业道德，属于不劳而获"搭便车"的不正当竞争行为，最终判令M公司赔偿T公司200万元。由此可见，虽然国家鼓励科技创新和技术进步，且技术本身是中立的，但当经营者将技术作为不正当竞争的手段或工具时，这种行为就具有了可罚性。

对于中国人工智能企业而言，通常采用自行采集、业务积累、爬虫抓取、数据购买、生态共享、算法生成、公开数据集这七种典型方式中的一种或几种获取自身所需的数据。依据数据的开放程度，涉及数据获取、使

2.杭州市中级人民法院,(2018)浙01民终7312号.

用的竞争行为会呈现不同的形态。

1.涉及完全公开数据的不正当竞争行为

完全公开数据的获取没有用户身份认证等事前限制，第三方数据使用者可以自由地获取数据。例如，数据使用者通过“网络爬虫”（Web Spider）抓取搜索引擎上的公开数据。而数据控制者可以通过设置robots协议拒绝爬虫访问。robots.txt文件本身默认值就是“允许”抓取，“不允许”抓取只是特例。当一个网站未设置robots.txt文件或robots.txt文件的内容为空时，则意味着该网站对于所有搜索引擎的网络爬虫都是开放的。在这种数据获取场景下，一方面，如果数据控制方通过robots协议对其它经营主体获取数据进行不合理的限制，就可能构成不正当竞争。

在“B公司与Q公司不正当纠纷案”[3]中，B公司一直在其相关网站的robots协议中排除Q公司的搜索引擎，用户使用Q公司的搜索引擎搜索到B公司的相关网站后，在点击访问时，会出现访问被阻断并跳转到B公司的搜索引擎网站的现象。对此，北京市高级人民法院认为，robots协议的初衷是为了指引搜索引擎的网络爬虫更有效地抓取对网络用户有用的信息，从而更好地促进信息共享。如果网站通过设置robots协议，使“允许”抓取成为特例，显然与robots协议的初衷背道而驰。根据《互联网搜索引擎服务自律公约》第八条的约定，robots协议对于通用搜索引擎抓取限制的设置应当具有行业公认合理的正当理由。B公司在缺乏合理、正当理由的情况下，以对网络搜索引擎经营主体区别对待的方式，限制Q公司的搜索引擎抓取其相关网站网页内容，影响Q公司搜索引擎的正常运行，这不仅会降低Q公司搜索引擎的用户满意度，损害Q公司的合法权益和相

3.北京市高级人民法院，（2017）京民终字第487号.

关消费者的利益，也会在客观上增强B公司搜索引擎的市场优势地位，妨碍正常的互联网竞争秩序，违反公平竞争原则，且违反诚实信用原则和公认的商业道德，构成《反不正当竞争法》第二条规定所指的不正当竞争行为。

另一方面，如果数据使用者使用网络爬虫从数据控制方抓取数据，并明显超过合理限度进行使用，也会构成不正当竞争。其中一种情形是数据使用者从数据控制方获取数据，并通过简单的算法处理后向用户提供与数据控制方同质的产品或服务，那么这种数据使用行为可被认为是"实质性替代"的不正当竞争。在"H公司诉B公司不正当竞争纠纷案"[4]中，法院认为B公司抓取涉案信息并不违反robots协议，但B公司在其产品中大量使用来自H公司用户的评论信息，已对H公司构成实质性替代，消减了H公司的竞争优势和交易机会，这种超出必要限度使用涉案信息的行为不仅损害了H公司的利益，也可能使得其他市场主体不愿再就信息的收集进行投入，破坏正常的产业生态，并对竞争秩序产生一定的负面影响。同时，这种超越边界的使用行为也可能会损害未来消费者的利益。就本案而言，如果获取信息投入者的利益不能得到有效保护，则必然使得进入这一领域的市场主体减少，消费者未来所能获知信息的渠道和数量亦将减少。B公司实施的是一种不正当竞争行为。

2.涉及相对公开数据的不正当竞争行为

相对公开数据的获取需要经过身份认证等事前授权。对该种数据的不正当获取方式，主要是指未经授权或者超出授权范围抓取用户数据。例如，数据使用者通过API接口认证获取数据，但是超越数据访问权限获取

4.上海知识产权法院，（2016）沪 73 民终 242号.

了授权范围之外的数据。在“W公司诉T公司不正当竞争纠纷案[5]”中，在T公司明确了解需要通过申请获得用户相关信息的接口权限，且合作终止后应当及时删除获取的用户信息的情况下，T公司在合作期间超出许可范围抓取并使用W公司用户职业信息、教育信息，并在合作终止后较长一段时间内仍然使用来自W公司用户信息；T公司的行为主观故意明显，行为违反了诚实信用的原则，违背了公认的商业道德，危害到W公司用户信息安全，损害了W公司的合法竞争利益，对W公司构成不正当竞争。大数据时代，通过API接口实现数据资源共享是目前企业之间合作的新模式，根据相关法律规定，经营者收集、利用用户信息应当遵循合法、正当、必要的原则并经用户同意。第三方通过API接口获取平台用户信息时应坚持“用户授权”+“平台授权”+“用户授权”的三重授权原则，即数据控制方在向第三方提供数据之前应先取得用户授权，而第三方应获得数据控制方的授权，并且在通过API接口获取用户信息并使用时需再次取得用户的授权，明确告知用户其使用的目的、方式和范围，尊重用户的知情权和自由选择权。

3、涉及不公开数据的不正当竞争行为

对于不公开数据而言，数据控制方一般对数据进行了技术措施保护。不公开数据如果具有秘密性、价值性和保密性，则属于商业秘密，对该类数据的非法获取是侵犯商业秘密的不正当竞争行为。在不公开数据类型领域中，典型的不正当数据获取行为是“黑客行为”，该种数据获取行为，不仅涉及不正当竞争，甚至可能构成刑事犯罪。例如，在“G公司与Y公司的数据不正当竞争纠纷案 ”[6][7]中，为了提高Y公司开发的智能公交APP“车来

5.北京知识产权法院，（2016）京 73 民终588号.
6.深圳市南山区人民法院，（2017）粤 0305 刑初第 153 号.
7.深圳市中级人民法院，（2017）粤 03 民初 822 号.

了"信息查询的准确度及在中国市场的用户量，保证公司更好的经营，Y公司相关人员利用网络爬虫软件攻破G公司APP的加密系统，大量爬取G公司开发的智能公交APP"酷米客"的实时数据，爬取的数据直接为Y公司所用，使该公司的智能公交APP"车来了"准确度提高。经评估：G公司因被非法侵入计算机信息系统所造成的直接经济损失为24.43万元人民币。深圳市南山区人民法院认定Y公司相关人员违反国家规定，采用其他技术手段，获取计算机信息系统中储存的数据，情节特别严重，其行为已构成非法获取计算机信息系统数据罪，并且，深圳市中级人民法院认为Y公司利用网络爬虫技术大量获取并且无偿使用原告G公司"酷米客"软件的实时公交信息数据的行为，实为一种"不劳而获"、"食人而肥"的行为，具有非法占用他人无形财产权益，破坏他人市场竞争优势，并为自己谋取竞争优势的主观故意，违反了诚实信用原则，扰乱了竞争秩序，构成不正当竞争行为。[8]

/PART 002

涉数据、算法等技术的新型不正当竞争行为的分类

有学者从反不正当竞争法学理视角进行分析，提出涉数据、算法等技术的新型不正当竞争行为可以被区分为两种基本类型：（一）非效能竞争风险类型；（二）阻碍竞争风险类型。

（一）非效能竞争风险类型

非效能竞争风险类型的不正当竞争行为是指：经营者利用数据、算法

8.李安,《人工智能时代数据竞争行为的法律边界》.

手段，显著妨碍消费者对产品、服务的优劣做出理性判断，从而使消费者丧失固有的在市场领域决定产品、服务优胜劣汰的裁判功能。非效能竞争风险类型不正当竞争行为包括：

1. 误导型不正当竞争行为

误导型不正当竞争行为系经营者滥用数据、算法手段，针对消费者制造"信息茧房"、"信息误导"效应，以致影响消费者做出理性决策。《征求意见稿》第15条描述的"经营者利用技术手段误导、欺骗用户修改、关闭、卸载、放弃使用其他经营者合法提供的网络产品或者服务"即属于典型的误导型不正当竞争行为。例如，上海市长宁区人民法院认为，合同双方通过各种手段让涉及某品牌的负面新闻在搜索引擎上不被社会公众所知晓或者不容易被社会公众所知晓，该"负面压制"条款的目的违背了诚实信用的基本法律原则，严重违反了《消费者权益保护法》和《反不正当竞争法》的基本原则，也将损害搜索引擎服务提供者的公信力，进而损害其商业价值，因而认定该"负面压制"条款无效。[9]

2. 侵犯型不正当竞争行为

侵犯型不正当竞争行为系经营者利用施加压力的方式，迫使消费者做出相应商业决策，从而严重损害消费者的自主决策权。《征求意见稿》第三章（第13-16条）所禁止的"利用技术手段实施妨碍干扰等不正当竞争行为"主要属于"侵犯型不正当竞争行为"范畴。如果经营者并不是从便利消费者与服务消费者的角度设定新型商业模式，而是试图通过强制改变用户消费习惯的方式来攫取商业暴利（例如，餐厅经营者强行以扫码点餐取代人工点餐），那么就涉嫌构成侵犯型不正当竞争行为。

9.严剑漪、王雨堃,《上海长宁法院判决确认两家公司合同中"负面压制"条款无效》，https://www.china-court.org/article/detail/2021/09/id/6259700.shtml.

（二）阻碍竞争风险类型

阻碍竞争风险类型的不正当竞争行为是指：经营者利用数据、算法手段，阻碍其他经营者提供高性价比的产品、服务，从而阻滞公平与有效的市场竞争机制的运行。

在反不正当竞争法律实践中，恶意不兼容行为属于典型的阻碍竞争风险类型的不正当竞争行为。例如，经营者利用搜索降权、排名后置、阻断流量等隐秘的数据、算法手段变相强迫平台内商家接受"二选一"的要求，进而阻碍竞争对手提供有效的服务，这种行为就构成阻碍竞争风险类型的不正当竞争行为。

在互联网市场竞争领域，如果经营者故意借助数据、算法手段实施一项不兼容行为，并且该项行为不符合诚实信用原则与公认的商业道德标准，那么该项行为就涉嫌构成恶意不兼容形态的不正当竞争行为。如果一个经营者不利用数据、算法手段实施不兼容行为，它将无法正常提供其网络产品与服务，抑或它的合法权益将受到侵害，或者，它将额外负担过于高昂的设施改造成本与维护成本，那么该经营者实施的不兼容行为就具有正当性与合法性。[10]

/PART 003
涉数据、算法的新型不正当竞争行为的法律规制

市场经济鼓励商业创新和公平、自由的竞争，以此实现优胜劣汰。但扰乱市场秩序的不正当竞争行为应当予以禁止。《中华人民共和国反不正当竞争法》（以下简称"**《反不正当竞争法》**"）第二章列举了几种常见的

10.翟巍，数据、算法驱动型不正当竞争行为的规制路径——兼评《禁止网络不正当竞争行为规定（公开征求意见稿）》

不正当竞争行为，但由于市场竞争行为方式具有多样性和可变性，法律不可能对所有类型的不正当竞争行为都预先作出规定。例如，语音指令作为人工智能中人机交互的一种方式，较之商品名称、企业名称、域名等出现得较晚，在类型方面亦存在一定差别，《反不正当竞争法》第六条并未予以列明。但是《反不正当竞争法》第六条的目的在于制止混淆行为，避免相关公众产生误认。因此，只要能够与该商品或服务及其提供者建立起特定的联系，且具有一定的影响，即应被纳入《反不正当竞争法》第六条所规定的权益保护范围之内。在"B公司与Z公司不正当竞争纠纷案"[11]中，北京海淀法院认为，B公司的语音指令已成为用户在使用其产品时必不可少且频繁出现的特定的语音指令，已与该产品的人机交互等功能和服务建立起密不可分的联系，且经过大量、广泛、形式多样的宣传推广，使得该语音指令具有较高的知名度和较大的影响力，与B公司及其产品建立起了明确、稳定的联系。B公司对其语音指令享有合法权益，应当受到反不正当竞争法第六条第四项的保护。Z公司使用与B公司完全相同的语音指令，在主观上难谓不具恶意，在客观上也极易导致相关公众产生混淆。Z公司的行为违反了反不正当竞争法第六条第四项的规定，对B公司构成不正当竞争。在上述案例中，语音指令的产生和识别都离不开人工智能算法和数据，当语音指令构成用于识别商品或服务的语音商标权时，其受到《反不正当竞争法》第六条第四项的保护，即经营者不得实施"其他足以引人误认为是他人商品或者与他人存在特定联系的混淆行为"。

随着互联网产业的蓬勃发展，出现了流量劫持、屏蔽广告、数据爬取、大数据杀熟等多种新型不正当竞争行为，而这些行为的实现往往借助了涉算法和数据的人工智能技术，传统的《反不正当竞争法》对于规制这

11.北京海淀法院，(2019)京0108民初63253号.

些新型不正当竞争行为存在滞后性问题。因此，2017年修订的《反不正当竞争法》在传统的六类不正当竞争行为，即仿冒混淆、商业贿赂、虚假宣传、侵犯商业秘密、有奖销售、商业诋毁之外，增设了第十二条"互联网专条"，强调对利用技术手段，通过影响用户选择或者其他方式，实施的妨碍、破坏其他经营者合法提供的网络产品或者服务正常运行的行为进行了法律规制，包括典型的插入链接、强制进行目标跳转的流量劫持行为，误导、欺骗、强迫用户修改、关闭、卸载其他经营者合法提供的网络产品或者服务的妨碍干扰行为，以及恶意对其他经营者合法提供的网络产品或者服务实施不兼容的恶意不兼容行为。2021年8月17日，国家市场监督管理总局（以下简称"**市监总局**"）发布了《关于<禁止网络不正当竞争行为规定（公开征求意见稿）>征求意见的通知》，包含多个针对涉数据、算法的新型不正当竞争行为进行细化的具体规定，例如，第十三条规定"经营者不得利用数据、算法等技术手段，通过影响用户选择或者其他方式，实施流量劫持、干扰、恶意不兼容等行为，妨碍、破坏其他经营者合法提供的网络产品或者服务正常运行"；第二十一条规定"经营者不得利用数据、算法等技术手段，通过收集、分析交易相对方的交易信息、浏览内容及次数、交易时使用的终端设备的品牌及价值等方式，对交易条件相同的交易相对方不合理地提供不同的交易信息，侵害交易相对方的知情权、选择权、公平交易权等，扰乱市场公平交易秩序。交易信息包括交易历史、支付意愿、消费习惯、个体偏好、支付能力、依赖程度、信用状况等"。

在涉数据、算法的新型不正当行为未落入《反不正当竞争法》第六条至第十二条的特别规定调整范畴时，可考虑适用《反不正当竞争法》第二条，该条款为"一般条款"，具有填补法律漏洞的作用。在具体案件中，对那些虽不属于《反不正当竞争法》第二章所列举，但确属违反诚实信用原则和公认的商业道德而具有不正当性的竞争行为，法院可以适用《反不正

当竞争法》第二条予以调整，以保障市场公平竞争。在"海带配额案"中，最高院明确了适用《反不正当竞争法》第二条应当同时具备以下条件（"三要件"标准）：1）法律对该种竞争行为未作出特别规定；2）其他经营者的合法权益因该竞争行为而受到了实际损害；3）该种竞争行为因确属违反诚实信用原则和公认的商业道德而具有不正当性——即违反"诚实信用原则和公认的商业道德"是证明该行为具有"不正当性"的核心要素。在"T公司与Q公司不正当竞争纠纷案"[12]中，最高人民法院认为，在市场经营活动中，相关行业协会或者自律组织为规范特定领域的竞争行为和维护竞争秩序，有时会结合其行业特点和竞争需求，在总结归纳其行业内竞争现象的基础上，以自律公约等形式制定行业内的从业规范，以约束行业内的企业行为或者为其提供行为指引。这些行业性规范常常反映和体现了行业内的公认商业道德和行为标准，可以成为人民法院发现和认定行业惯常行为标准和公认商业道德的重要渊源之一。由此，最高人民法院认可将《互联网终端软件服务行业自律公约》第18条、第19条作为认定互联网行业惯常行为标准和公认商业道德的参考依据。最高人民法院于2022年3月16日发布的《最高人民法院关于适用〈中华人民共和国反不正当竞争法〉若干问题的解释》第三条也规定，特定商业领域普遍遵循和认可的行为规范，人民法院可以认定为《反不正当竞争法》第二条规定的"商业道德"。

除上述提及的《反不正当竞争法》《消费者权益保护法》《最高人民法院关于适用〈中华人民共和国反不正当竞争法〉若干问题的解释》等法律法规及司法解释外，《民法典》《网络安全法》《数据安全法》《个人信息保护法》《互联网信息服务算法推荐管理规定》《广告法》等相关法

12.最高人民法院，（2013）民三终字第5号.

律法规也通过多个条款对涉数据、算法的不正当竞争行为进行了规制，例如，《数据安全法》第五十一条规定，“窃取或者以其他非法方式获取数据，开展数据处理活动排除、限制竞争，或者损害个人、组织合法权益的，依照有关法律、行政法规的规定处罚”；《互联网信息服务算法推荐管理规定》第十五条规定，“算法推荐服务提供者不得利用算法对其他互联网信息服务提供者进行不合理限制，或者妨碍、破坏其合法提供的互联网信息服务正常运行，实施垄断和不正当竞争行为”。这些法律法规均明确不得利用算法实施不正当竞争行为，要使用合法合规的方式进行数据获取和使用，充分保护数据及其相关主体的合法权益。

/PART 004
对经营者开展涉数据、算法的商业竞争行为的实践建议

自由竞争能够确保市场资源优化配置，但经营者开展涉数据、算法的商业竞争行为时应当遵守诚实信用原则和公认的商业道德，维护公平竞争的市场秩序，保护经营者和消费者的合法权益，不得损害国家利益和社会公共利益。笔者通过对相关司法实践的总结，谨在此给出一些实践建议供参考：

第一，由于目前人工智能产业中的相关行业规范、标准及自治协议仍处在发展阶段，经营者应当把握机会，在不违反法律原则和规则的前提下，在开发新商业模式的同时积极推进和建立行业规范及自治协议，并得到行业内经营者的广泛签署，使它们成为《反不正当竞争法》一般条款所规定之“商业道德”的具体化、特定化内容，并在涉数据、算法等人工智能技术的应用过程中严格遵循这些行业规范和自治协议，争取将其作为评价行为具有“正当性”的依据。

第二，经营者作为数据控制方若在robots协议中设置限制条件需注意

合理性和正当性。目前，由于robots协议可能会被一些行业巨头利用作为垄断的工具，包括互联网工程任务组（IETF, Internet Engineering Task Force）在内的一些重要国际组织仍拒绝采纳robots协议作为行业标准。如果经营者作为数据控制者希望通过设置robots协议限制网络爬虫抓取数据，应当具有合理、正当的理由，例如，出于保护受访网站内部信息或敏感信息的需要；出于维护受访网站正常运行的需要；出于保护社会公共利益的需要等；并且，不得影响用户的自主选择权。

第三，经营者作为数据获取方在获取不同开放程度的数据时需遵循"合法、正当、必要"原则。从上篇中的案例可以看出，针对不同开放程度的数据，经营者或多或少会使用算法等技术手段进行不同方式的数据获取，帮助其实现商业竞争目的。针对完全公开数据，若数据控制方设置了不合理的robots协议，经营者可以遵循"协商-通知"原则处理，向数据控制方提出书面修改robots协议、准许数据获取方获取数据的请求。若数据控制方在合理的期限内未书面、明确地提出其拒绝修改robots协议的合理理由的，或者数据获取方认为理由不成立的，可以请求相关执行机构或行业协会先行调解和裁决，或采取诸如诉讼、申请行为保全等法律措施予以解决。经营者在未违反robots协议的前提下通过搜索引擎抓取数据，也需遵循"合法、正当、必要"原则，采取对数据控制者损害最小的措施，例如，可考虑设置指向数据源的链接，符合"协商—通知"规则，以避免对具有竞争关系的数据控制者的业务构成实质性替代。针对相对公开数据，第三方经营者通过API接口获取平台用户信息时应坚持"用户授权"+"平台授权"+"用户授权"的三重授权原则。避免未经授权采用技术手段非法获取数据，否则不仅可能涉及不正当竞争行为，侵犯商业秘密，甚至还可能触犯刑法，构成非法获取计算机信息系统数据罪。

第四，经营者在使用数据时可考虑结合司法案例从以下几个方面进行"正当性"论证，（1）数据是否具有商业价值，能否给经营者带来竞争优

势；（2）数据获取的难易程度和成本付出；（3）对数据的获取及利用是否违法、违背商业道德或损害社会公共利益；（4）竞争对手使用的方式和范围。如果数据使用者从数据控制方获取数据后，通过算法的深度加工和挖掘，向市场提供了类似于数据控制方的、但更为优质的产品或服务，那么这种数据使用行为具有经济合理性，不易被认定为不正当竞争。此外，在人工智能领域，用户的隐私权也是一个非常重要的课题。在人工智能时代，对用户数据的使用，一方面要规范数据使用技术，加强数据匿名化处理；另一方面要明确数据的使用用途，不得侵犯个人隐私。

第五，经营者应增强人工智能技术应用流程的可控性与透明度，最小化算法黑箱特性可能带来的负面影响和损害后果。在现行法律体系之下，人工智能算法和数据常被视为核心的竞争性资源而作为商业秘密进行保护。但是，这些人工智能技术会对用户的消费决策产生较大影响，虽然消费者并不一定具有足够的专业知识或者并不想了解这些技术意义，经营者仍应在广告宣传、信息过滤、商业评价等方面向消费者做出关于技术工作原理的充分说明，将自主权和选择权交还消费者，而不能强制或暗中为消费者作出决定，以确保算法的可解释性与可验证性。

/PART 005
结语

反不正当竞争目的在于维护公平的市场竞争秩序，在2020年至2023年，全国法院共审理不正当竞争民事案件21,893件[在威科先行数据库以"民事""不正当竞争纠纷"和"裁判日期：最近三年"为条件进行检索，截止时间为2023年12月5日。]，出现了广告屏蔽、数据抓取、流量劫持等受到社会高度关注的新型不正当竞争民事案件。随着人工智能技术的快速发展和消费者需求的改变，经营者必然会不断改进商业模式，市场竞争秩序

将逐渐被人工智能技术影响和改变，可以预见未来人工智能技术将对反不正当竞争法提出诸多挑战。经营者在谋求正当商业利益和不损害他人合法权益的前提下，提供尽可能便利消费者选择或者更好满足消费需求的中立性技术工具或者手段，非但不会受到法律禁止，而且还会得到市场激励。人工智能时代的商业竞争行为应以法律为边界，树立正确的价值观，“倡导向上向善，抵制逐利作恶”，合理规制涉数据、算法的竞争行为，确保公平、透明，为人工智能产业的发展营造一个公平自由的市场竞争环境。

（陈茜对本文亦有贡献）

王红燕
合伙人
知识产权部
杭州办公室
+86 571 5662 3968
gracewang@zhonglun.com

人工智能服务提供者的过错责任初探

ARTICLE BY 赵刚 高敏

在可见的未来，人工智能技术必将持续、深入发展，迈进大规模商用阶段。但人工智能的前景虽令人憧憬，也存在一定法律风险。本文立足于人工智能发展现状，对应用层面的人工智能服务提供者的过错责任承担问题进行探究，以供感兴趣的读者参考。

/PART 001

人工智能的法律主体地位

目前，对于人工智能产品是否具有法律主体地位存在一定争议。部分学者认为可以依据法人的相关规定类推赋予人工智能产品拟制法律人格，由其享有权利，并在人工智能产品给他人造成损害时，由其承担相应的民事或刑事责任。我们可以想见，如果将人工智能产品视为独立的法律人格主体，则可能将明显减轻人工智能服务提供者及人工智能产品生产者的法律责任。

但是鉴于人工智能目前还正处于由弱向强的过渡转变阶段，弱人工智能还不具备成为拟制人的条件。并且人工智能产品的相关行为是人类意志的结果，人工智能本质上是科技领域的人类智力劳动成果，其存在是为了服务于人类繁衍、生存、发展的终极目的，不管人工智能发展到何种高级阶段，其作为人类工具的属性不会改变，因此其仍可属于物的范围，人工智能不具有法律主体地位的观点目前也仍为主流。因其不能独立地做出意思表示，没有民事行为能力和民事权利能力，不具有承担责任的能力，因其产生的相关法律责任仍需由其他主体承担。

在处理人工智能侵权纠纷的司法实践中，法院对于人工智能侵权责任主体的确定仍持保守审慎态度，其认为人工智能属于平台服务的部分内容，对应的网络服务提供者在一定条件下需要为其服务活动承担责任。在最高人民法院发布的人格权司法保护典型民事案例中[1]，法院认为某AI软件擅自将某公众人物的的姓名、肖像、人格特点等综合而成的整体形象投射到AI角色上，形成了该公众人物的虚拟形象，属于对整体人格形象的使

1.最高人民法院《民法典颁布后人格权司法保护典型民事案例》之四

用。同时某AI软件运营者通过算法应用，将该角色开放给众多用户，用户可以与该AI角色设定身份关系、设定任意相互称谓、通过制作素材"调教"角色，从而形成与该公众人物真实互动的体验，对于案件的上述功能设置还涉及自然人的人格自由和人格尊严。虽然具体图文由用户上传，但某AI软件运营者的产品设计和对算法的应用实际上鼓励、组织了用户的上传行为，直接决定了软件核心功能的实现，因此不再只是中立的技术服务提供者，最终判决该AI软件运营者作为内容服务提供者承担侵权责任。

根据《生成式人工智能服务管理办法》（以下简称"**《办法》**"）第九条规定，"提供者应当依法承担网络信息内容生产者责任，履行网络信息安全义务。涉及个人信息的，依法承担个人信息处理者责任，履行个人信息保护义务"，人工智能服务提供者需要承担人工智能的内容生产者责任，人工智能服务提供者包括研发、利用生成式人工智能产品的技术研发商、应用开发商与提供API接口等接入服务的提供商等。该《办法》已于2023年8月15日正式生效，虽相较《生成式人工智能服务管理办法（征求意见稿）》第五条而言有一定变动，但是总体上反应了针对人工智能的立法趋势与考量。

/PART 002
人工智能服务提供者的法律地位

我国《信息网络传播权保护条例》将网络服务提供者大致分为四个类型，分别为提供网络接入、传输中介的网络服务提供者；提供信息自动缓存的网络服务提供者；提供网络存储服务的网络服务提供者；提供网络信息定位（即搜索链接服务）的网络服务提供者。与人工智能服务提供者较为近似的为提供网络接入、传输中介的网络服务提供者及提供网络信息定位的网络服务提供者。

首先，《办法》第五条要求提供API接口接入服务的人工智能服务提供者承担人工智能产品的内容生产者责任。我们理解，该责任义务重于网络接入、传输中介的网络服务提供者所需承担的法律责任。网络接入、传输中介的网络服务提供者需承担一般性的"注意义务"，一般不包括承担事前的审查义务，但在知悉侵权行为存在时应及时采取删除、屏蔽、断开链接等必要措施。

其次，信息定位服务与人工智能服务均根据用户输入的关键词需求，通过算法匹配相应结果，但信息定位服务还需用户点击搜索结果链接跳转第三方网站获取内容，而人工智能服务则直接生成并向用户提供内容。信息定位服务通常并不涉及直接参与第三方网站侵权内容的生产，亦无法控制第三方网站侵权内容的产生，但人工智能服务对于侵权内容的产生具有控制力，因此我们倾向于认为，人工智能服务不应归属于信息定位服务范畴。

因此我们理解，人工智能服务提供者与我国现有的网络服务提供者类型暂未全面契合，属于新型的网络服务提供者类型，其承担人工智能侵权责任标准和范围目前需要进一步明确界定。

/PART 003
人工智能服务提供者的归责原则

人工智能可能导致的侵权风险复杂多样，主要包括数据合规风险、生成内容滥用风险、算法滥用风险、隐私保护风险等法律风险。目前我国对其直接进行规制的相关法律法规有待进一步完善，针对人工智能服务提供者的侵权责任可依据我国侵权责任归责原则确定。我国规定的侵权责任归责原则包括过错责任原则、无过错责任原则和公平责任原则，过错责任原则是侵权责任中最基本、最主要的归责原则。

虽然人工智能因算法复杂并且由人工智能服务提供者所掌握，受害者证明人工智能服务提供者对侵权行为的产生具有过错、侵权行为与损害结果具有因果关系等构成要件具有较大难度；但如将人工智能产品适用产品责任进行归责，追究人工智能服务提供者的无过错责任目前也较难实现。因此针对人工智能侵权行为，要求人工智能服务提供者承担侵权责任仍需判定其具有主观过错，考察人工智能服务提供者主观对侵权行为是否知情。

通常而言，网络服务提供者不具有一般性审查义务，因此对于侵权内容的出现不会直接被认定为存在过错；但在其明知或应当知道侵权内容后，则网络服务提供者负有及时采取必要措施的义务，否则认定其存在过错。"明知"要求网络服务提供者明确知晓侵权行为的存在，比如人工智能服务提供者提供人工智能服务的用途仅为生成特定的侵权内容，积极追求侵权结果的发生，否则一般较难以以"明知"标准认定过错。"应知"则通过规定人工智能服务提供者一定的注意义务，在其具备相应注意能力的情况下，应当或者能够认识到侵权行为的发生。我们理解，《办法》等相关法律法规对于人工智能服务提供者的规制着眼于事前预防，侧重对算法的直接规制，从而判断人工智能服务提供者对于算法生成侵权内容是否具有过错，与传统的网络服务提供者"通知-删除"义务有较大的区别。

/PART 004

人工智能服务提供者侵权责任的认定

对于人工智能服务提供者的过错判断，可参照《办法》规定的人工智能服务提供者所需履行的部分义务，包括安全评估及算法备案义务、用户信息保护义务、数据来源合法义务及内容合规义务等，如人工智能服务提供者违反相应的义务，则可能具有相应的过错。

（一）监管机制相关义务

首先，《办法》第十七条规定，“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，应当按照国家有关规定开展安全评估，并按照《互联网信息服务算法推荐管理规定》履行算法备案和变更、注销备案手续”，因此人工智能服务提供者具有对人工智能产品进行安全评估与算法备案义务，在其对外提供服务的产品的显著位置标明备案编号，并提供公示信息链接，便于监管部门通过算法备案了解相关人工智能服务的算法属性，并根据安全评估实现对相关人工智能服务信息的了解。

其次，算法的实施会带来“算法黑箱”，算法决策的规则通常会被算法开发者所隐蔽，缺乏透明性。此前国家互联网信息办公室等联合发布的《互联网信息服务算法推荐管理规定》[2]针对算法推荐提出了公开透明原则。《办法》第十九条[3]亦规定了人工智能服务提供者具有提高算法透明度和可解释性的义务，以应对相关监管部门的“算法服务检查”，履行披露义务。

最后，与其他类型的网络服务提供者一致，人工智能服务提供者亦负有设置便捷的用户申诉和公众投诉、举报入口的义务，公布处理流程和反馈时限，方便公众或潜在被侵权人向人工智能服务提供者反馈意见或投诉。

（二）用户相关义务

《办法》第九条、十一条要求人工智能服务提供者承担用户个人信息

2.《互联网信息服务算法推荐管理规定》第十六条规定，“算法推荐服务提供者应当以显著方式告知用户其提供算法推荐服务的情况，并以适当方式公示算法推荐服务的基本原理、目的意图和主要运行机制等”。

3.《生成式人工智能服务管理办法》第十九条规定，“有关主管部门依据职责对生成式人工智能服务开展监督检查，提供者应当依法予以配合，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，并提供必要的技术、数据等支持和协助”。

管理义务，第十条要求其承担用户防沉迷义务，第十四条要求其对用户不法行为进行管理的义务。《生成式人工智能服务管理办法（征求意见稿）》第十八条曾要求其承担知道用户理性使用人工智能服务的义务，避免诱导用户实施不法行为，《办法》虽将该条款予以删除，但也可反映一定的立法趋势。

（三）算法训练相关义务

算法是人工智能的核心，但人工智能服务提供者在利用算法技术提供服务时，也暴露出算法歧视、算法滥用在内的诸多风险，具体表现包括人工智能生成的虚假信息未作显著标识而夸大传播、人工智能仿制虚假内容造成混淆、人工智能扎堆推送负面信息诱发社会恐慌等。因此《办法》针对算法内容规定了事前、事中、事后全链条的义务。

首先，《办法》第七条[4]规定，人工智能服务提供者对于算法生成、优化的训练数据来源具有确认合法性的义务，应确保训练数据符合法律法规规定、不包含侵害知识产权的内容、个人信息取得同意、数据真实、准确、客观、多样等。因此人工智能服务提供者在数据来源收集过程中，应取得数据主体的知情且同意，并且依据最小必要原则，确保数据来源的合法性以及合理性。

其次，算法推荐机制是形成算法歧视的手段，一定程度上也是对算法歧视的强化，也极易对个体形成“信息茧房”，剥夺了对于信息的选择空间，因此《办法》要求人工智能服务提供者承担显著标识义务，对人工智

4.《生成式人工智能服务管理办法(征求意见稿)》第七条规定，“生成式人工智能服务提供者（以下称提供者）应当依法开展预训练、优化训练等训练数据处理活动，遵守以下规定：（一）使用具有合法来源的数据和基础模型；（二）涉及知识产权的，不得侵害他人依法享有的知识产权；（三）涉及个人信息的，应当取得个人同意或者符合法律、行政法规规定的其他情形；（四）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；（五）《中华人民共和国网络安全法》《中华人民共和国数据安全法》《中华人民共和国个人信息保护法》等法律、行政法规的其他有关规定和有关主管部门的相关监管要求。”

能生成内容及算法推荐内容明确标注，以保障用户的知情权。

最后，人工智能服务提供者负有算法纠偏义务，应定期审核、评估、验证生成合成类算法机制机理，在发现违法内容时采取停止生成、停止传输、消除等处置侵权内容，并应通过模型优化训练等方式进行整改。我们理解，该条实际对人工智能服务提供者施加了更为严格的注意义务。

（四）内容管理相关义务

《办法》第四条对于人工智能产生内容作出了相应的规定，即不得包含颠覆国家政权、推翻社会主义制度，煽动分裂国家、破坏国家统一，宣扬恐怖主义、极端主义，宣扬民族仇恨、民族歧视，暴力、淫秽色情信息，虚假信息，以及可能扰乱经济秩序和社会秩序的内容，不得出现种族、民族、信仰、国别、地域、性别、年龄、职业等歧视，不得生成虚假信息等。

该条款实际赋予人工智能服务提供者采取手段措施履行事前审查义务，要求其采取技术或者人工方式对用户输入数据和合成结果进行审核，以控制人工智能服务的算法运行和生成结果。虽然人工智能数据极为庞大，运营成本较高，人工智能服务提供者认为人工智能服务是种实质性非侵权用途的中立技术，要求其事前全面审查不具有可行性，但是在涉及推翻国家政权、民族亲属等侵权内容生成上，人工智能服务提供者却可较好地避免该侵权内容的产生，因此对于人工智能所产生其他类型的侵权内容，人工智能服务提供者亦应达到相应的注意义务标准，防止该类侵权内容的产生。

但是人工智能服务主体是否有能力做到全流程的审查监管，从而避免用户利用其服务产生侵权的内容，尚存在一定疑问，需根据行业的通常技术水平与技术水平的发展程度而确定。如果人工智能服务提供者可实现全流程的审查和监管，该行为亦存在侵害用户隐私权利的法律风险。

/PART 005

结语

随着人工智能服务的迅速发展，人工智能可能在不久的未来全方位融入人们的日常生活，其所带来的法律风险影响亦将日渐突显，政府部门的监管要求也将日益细化。因此为减少侵权风险，人工智能服务提供者应在创新发展人工智能技术的同时，采取相应技术手段减少侵权内容的产生，并积极履行对人工智能产品生成内容符合监管要求的事前审核义务，及时清除侵权内容，从而降低承担人工智能侵权责任的法律风险。

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

A

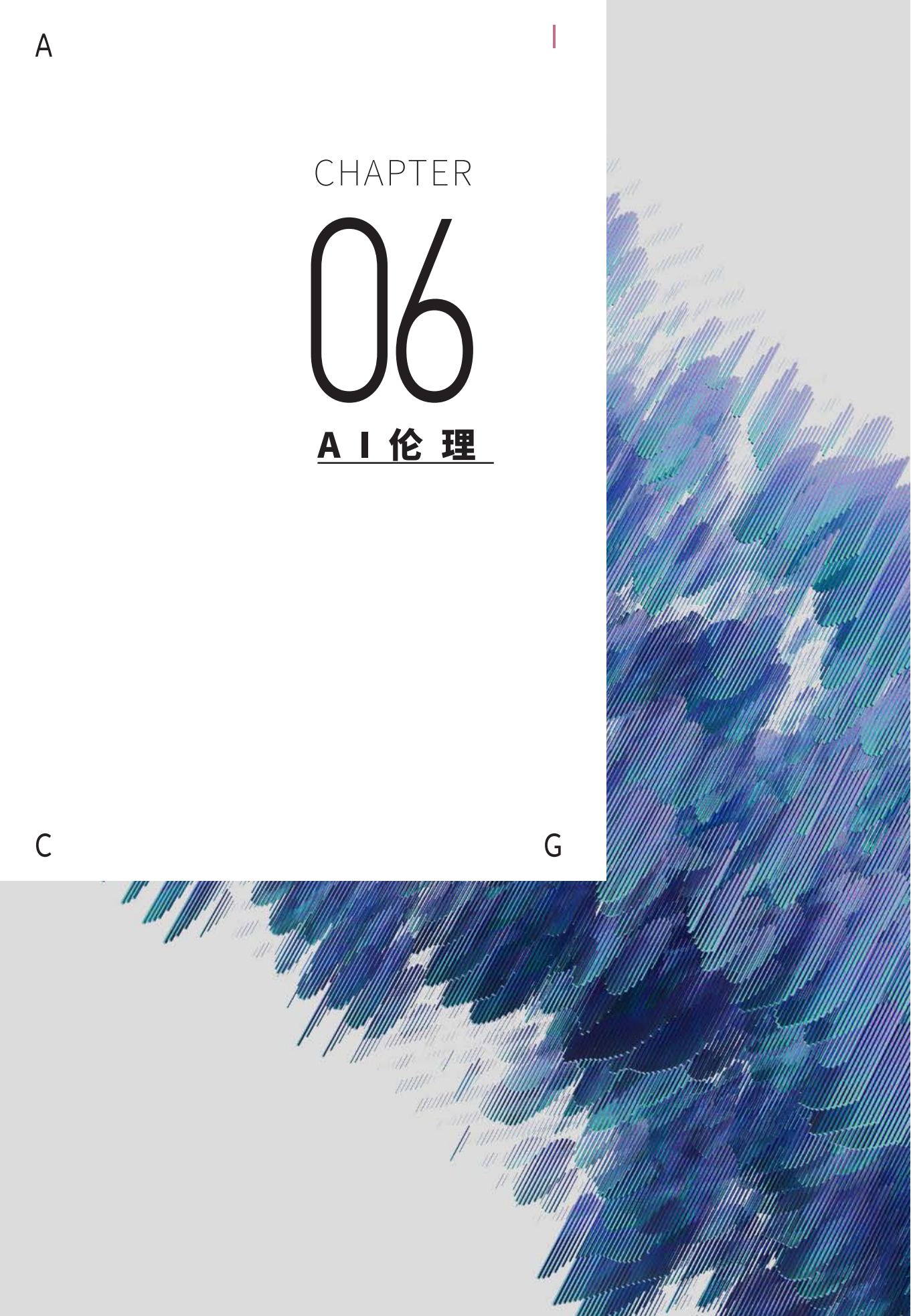
C

G

CHAPTER

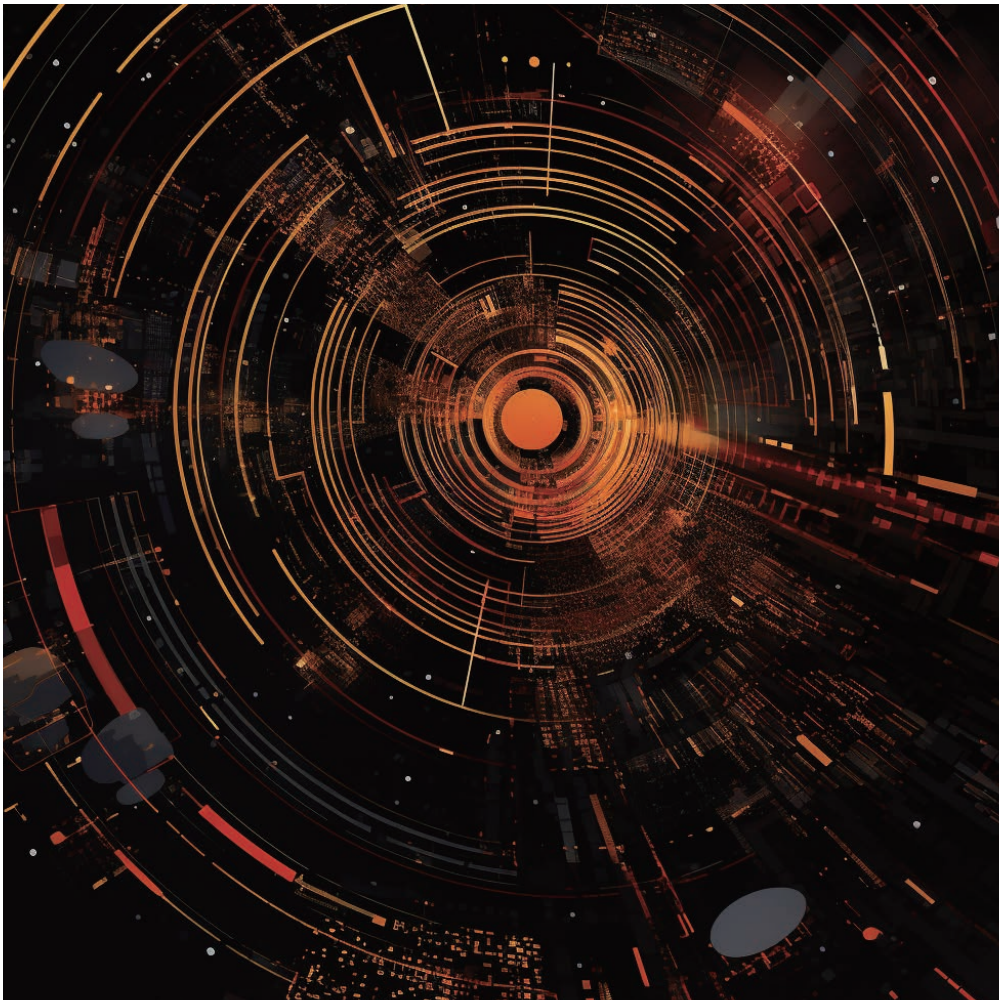
06

AI伦理



生成式AI合规探讨系列

——生成式AI伦理治理



ARTICLE BY 樊晓娟 印磊 黄海逸

生成式AI在踏入高速的发展轨道之时，其伦理治理的问题也备受社会各界关注。本文在分析生成式AI所带来的伦理问题基础之上，系统地阐述我国伦理治理体系。

/PART 001

伦理问题

当下，AI应用所引发的伦理问题已经引起了全球范围的广泛关注。2022年年初，联合国教科文组织出版了《人工智能伦理问题建议书》。该建议书提及："人工智能系统引发了新型伦理问题，包括但不限于其对决策、就业和劳动、社交、卫生保健、教育、媒体、信息获取、数字鸿沟、个人数据和消费者保护、环境、民主、法治、安全和治安、双重用途、人权和基本自由（包括表达自由、隐私和非歧视）的影响。此外，人工智能算法可能复制和加深现有的偏见，从而加剧已有的各种形式歧视、偏见和成见，由此产生新的伦理挑战。"

作为AI技术中的重要组成部分，生成式AI所涉及的伦理问题包括但不限于如下方面：社会交往的缺失、歧视和虚假信息以及劳动力替代。

/PART 002

伦理治理的原则和立场

中国政府对于人工智能伦理治理的立场和原则非常明确且坚定。2022年11月16日，中方向联合国《特定常规武器公约》缔约国大会提交了《中国关于加强人工智能伦理治理的立场文件》（以下简称"**《立场文件》**"）。《立场文件》指出"中国始终致力于在人工智能领域构建人类命运共同体，积极倡导'以人为本'和'智能向善'理念，主张增进各国对人工智能伦理问题的理解，确保人工智能安全、可靠、可控，更好赋能全球可持续发展，增进全人类共同福祉。"在监管方面，《立场文件》提出了多项重要主张，包括：坚持伦理先行；建议建立和完善人工智能伦理准则、规范及问责机制；建议建立和完善人工智能伦理的规范和法律政策体系；以及增强底线思维和风险意识等。

2022年3月20日《中共中央办公厅 国务院办公厅关于加强科技伦理治理

的意见》，明确规定科技伦理的原则为："增进人类福祉、尊重生命权利、坚持公平公正、合理控制风险、保持公开透明。"

/PART 003

伦理治理的立法情况

2016年，《涉及人的生物医学研究伦理审查办法》（以下简称"**《办法》**"）发布，该《办法》针对各级各类医疗卫生机构开展涉及人的生物医学研究进行伦理审查工作。

为适应人工智能的科技发展，在《办法》的基础上，2023年2月18日，国家卫生健康委员会等印发了《涉及人的生命科学和医学研究伦理审查办法》（以下简称"**《审查办法》**"）。《审查办法》将适用范围从《办法》的医学伦理审查扩展到了以人为受试者或者使用人（统称研究参与者）的生物样本、信息数据（包括健康记录、行为等）所开展的研究活动，并对其进行伦理审查。

2023年9月7日，科学技术部进一步发布了《科技伦理审查办法（试行）》（以下简称"**《试行办法》**"），旨在进一步完善科技活动的伦理审查，弥补医学伦理审查范围以外的科技活动的伦理审查规定的空白。

以下是三份文件适用范围和适用对象的比较：

法律依据	生效日期	适用对象	适用范围
《涉及人的生物医学研究伦理审查办法》	2016年12月1日	各级各类医疗卫生机构	涉及人的生物医学研究包括以下活动： （一）采用现代物理学、化学、生物学、中医药学和心理学等方法对人的生理、心理行为、病理现象、疾病病因和发病机制，以及疾病的预防、诊断、治疗和康复进行研究的活动； （二）医学新技术或者医疗新产品在人体上进行试验研究的活动； （三）采用流行病学、社会学、心理学等方法收集、记录、使用、报告或者储存有关人的样本、医疗记录、行为等科学研究资料的活动。
《涉及人的生命科学和医学研究伦理审查办法》	2023年2月18日	在中华人民共和国境内的医疗卫生机构、高等学校、科研院所等	以人为受试者或者使用人（统称研究参与者）的生物样本、信息数据（包括健康记录、行为等）开展的以下研究活动： （一）采用物理学、化学、生物学、中医药学等方法对人的生殖、生长、发育、衰老等进行研究的活动； （二）采用物理学、化学、生物学、中医药学、心理学等方法对人的生理、心理行为、病理现象、疾病病因和发病机制，以及疾病的预防、诊断、治疗和康复等进行研究的活动； （三）采用新技术或者新产品在人体上进行试验研究的活动； （四）采用流行病学、社会学、心理学等方法收集、记录、使用、报告或者储存有关人的涉及生命科学和医学问题的生物样本、信息数据（包括健康记录、行为等）等科学研究资料的活动。

法律依据	生效日期	适用对象	适用范围
《科技伦理审查办法（试行）》	2023年12月1日	高等学校、科研机构、医疗卫生机构、企业等	（一）涉及以人为研究参与者的科技活动，包括以人为测试、调查、观察等研究活动的对象，以及利用人类生物样本、个人信息数据等的科技活动； （二）涉及实验动物的科技活动； （三）不直接涉及人或实验动物，但可能在生命健康、生态环境、公共秩序、可持续发展等方面带来伦理风险挑战的科技活动； （四）依据法律、行政法规和国家有关规定需进行科技伦理审查的其他科技活动。

相应规定适用范围内的活动，应当通过伦理审查，适用范围发生竞合的，应当同时通过不同的伦理审查，未通过伦理审查的相应活动，可能面临行政处罚和刑事责任。

/PART 004

科学伦理审查

现阶段，中国关于伦理治理的主要方法为伦理审查。《试行办法》的发布，把伦理审查的范围从直接以人为研究对象的科技活动扩展到所有存在伦理风险的科技活动，弥补了医学伦理审查范围以外科技活动的伦理审查规定的空白。

（一）科学伦理审查责任主体

根据《试行办法》第四条的规定，高等学校、科研机构、医疗卫生机构、企业等是科技伦理审查管理的责任主体。从事生命科学、医学、人工智能等科

技活动的单位，研究内容涉及科技伦理敏感领域的，应设立科技伦理（审查）委员会。其他有伦理审查需求的单位可根据实际情况设立科技伦理（审查）委员会。

（二）科学伦理审查流程

适用范围内的科技活动必须通过科技审查后方可开展。一般审查由科技活动负责人向本单位科技伦理（审查）委员会申请伦理审查，科技伦理（审查）委员会决定是否受理并给予书面通知。

《试行办法》还建立了需要开展专家复核的科技活动清单制度，对可能产生较大伦理风险挑战的新兴科技活动实施清单管理，属于清单内的科技需要经过地方或行业主管部门专家复核程序方可开展。

（三）审查重点

根据《试行办法》第十五条的规定，科技伦理审查的标准和重点包括：科技人员资质、科学价值和社会价值、招募方案的公平合理、信息处理的合规性、科技活动研究参与者的权利保护、实验动物福利；数据处理和算法治理的合理性，以及利益冲突和管理方案合理性等等。

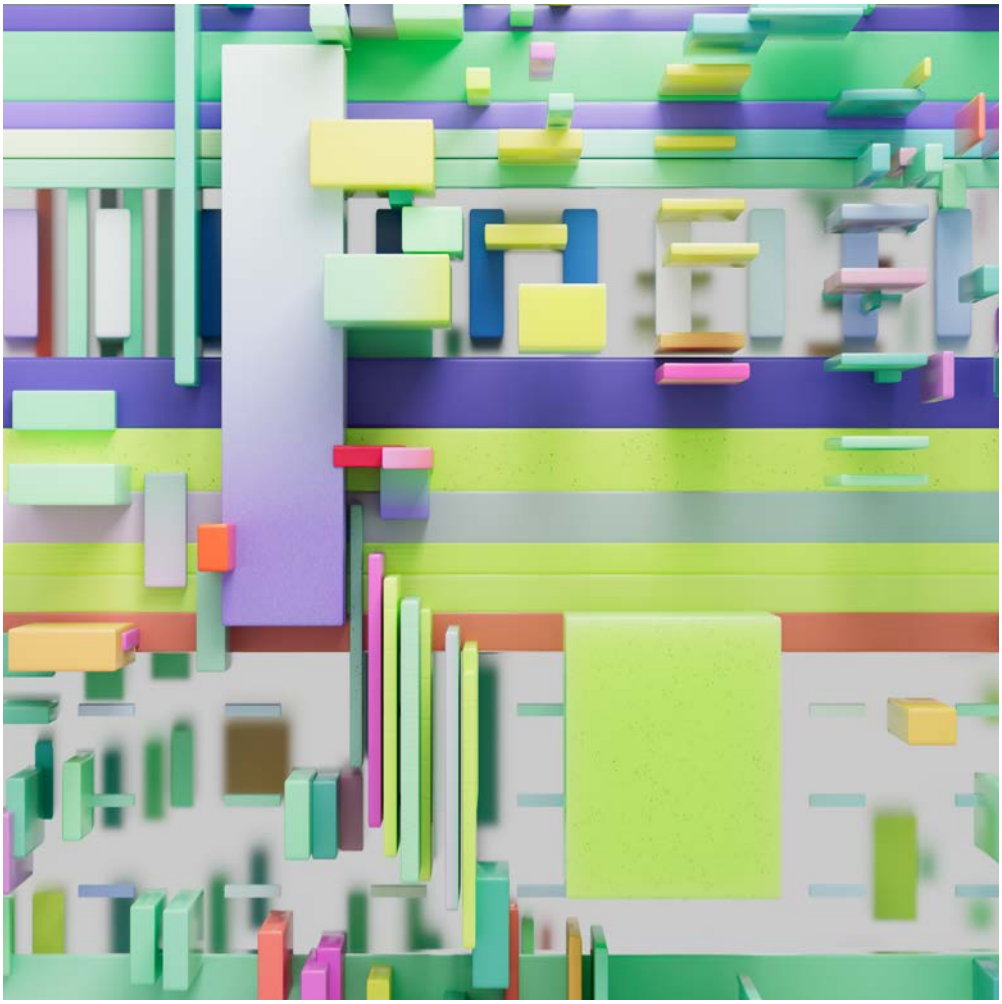
/PART 005
结语

AIGC是一项具有划时代意义的技术，也是当下飞速发展的人工智能影响人类社会政治经济等方面的典型例子。然而，AIGC也仍然存在许多法律问题并可能阻碍其广泛使用。回望过去，工业革命期间，人们也对自动化和蒸汽机等新技术产生了焦虑，但是工业革命所带来的成就促使人类实现了生产力的飞跃。当下，在看到AI技术发展所带来机遇的同时，我们也应当保持一个审慎的

态度，对AI技术可能的风险进行充分的考虑，并采取必要的防范措施来确保技术的安全使用。

樊晓娟
合伙人
私募基金与资管部
上海办公室
+86 21 6061 3669
fanxiaojuan@zhonglun.com

人工智能生成内容的合规监管与伦理道德小议



ARTICLE BY 顾萍 詹凯维

随着全球各大公司纷纷涌入人工智能生成内容（"AIGC"）行业，"一键生成"、"改头换面"等AIGC技术也随之普及。人们逐渐对与AIGC相伴而生的虚假信息、侵犯权益、道德伦理、偏见歧视等问题产生担忧。为进一步规范AIGC健康发展和规范应用，在充分吸收了各界意见后，国家互联网信息办公室（"国家网信办"）联合国家发展和改革委员会、教育部、科学技术部（"科技部"）、工业和信息化部（"工信部"）、公安部和国家广播电视总局六部门于2023年7月10日发布了《生成式人工智能服务管理暂行办法》（"《AI服务办法》"），该办法于2023年8月15日生效。《AI服务办法》是我国第一部关于AIGC的法律规范，旨在规范利用生成式人工智能（"AI"）技术向中华人民共和国境内公众提供生成文本、图片、音频、视频等内容的服务。

根据《AI服务办法》第四条，AIGC应当符合"遵守法律、行政法规"和"尊重社会公德和伦理道德"等两方面主要要求，包括不得生成违反主流价值观、虚假有害等法律、行政法规禁止的内容；应当采取有效措施防止不合理歧视；尊重知识产权，不得进行不正当竞争；尊重他人合法权益；提高内容准确性和可靠性等等。有鉴于此，我们尝试根据现行法律法规，梳理出上述法律法规和伦理道德的基本内容，并提出提出部分针对性建议如下，供相关企业参考。

/PART 001

国家法律、行政法规合规体系梳理

《AI服务办法》所提到的"法律、行政法规"，包括下述四部法律，未明确提到具体的行政法规。特别地，《AI服务办法》还提到了以下两份部门规章，分别为《互联网信息服务算法推荐管理规定》（"**《算法推荐规定》**"），《互联网信息服务深度合成管理规定》（"**《深度合成规定》**"）。我们将该等法律、部门规章的基本信息和主要内容总结如下。

1.《中华人民共和国网络安全法》

2016年11月7日，全国人民代表大会常务委员会（"**全国人大常委会**"）通过《网络安全法》（主席令第53号），该法律于2017年6月1日生效，系网络安全和监管领域的基础性法律。《网络安全法》确立依法使用网络，遵守公序良俗，不得危害网络安全，不得利用网络从事违法犯罪行为，并且对网络信息的内容和传播负有一定的管理义务等义务和责任。《网络安全法》设立网络安全等级保护制度，网络安全监测预警和信息通报制度，以及要求运营者设立网络信息安全投诉、举报制度，用户信息保护制度等。

2.《中华人民共和国数据安全法》

2021年6月10日，全国人大常委会通过《数据安全法》（主席令第84号），该法律于2021年9月1日生效，系数据安全和监管领域的基础性法律。《数据安全法》要求数据处理者在开展数据处理活动时，应当依法依规，尊重公序良俗，诚实守信，履行数据安全保护义务，承担社会责任，不得危害国家安全、公共利益，不得损害个人、组织的合法权益。《数据安全法》建立数据分类分级保护制度，数据安全风险评估、报告、信息共享、监测预警机制，数据安全应急处置机制，数据安全审查制度。就数据保护义务而言，《数据安全法》要求数据处理者建立健全全流程数据安全管理制度，建立数据安全风险和事件的监测和应对机制，定期风险评估机制，建立数据出境安全评估机制，数

据交易基本制度等。

3.《中华人民共和国个人信息保护法》

2021年8月20日，全国人大常委会通过《个人信息保护法》（主席令第91号），该法律于2021年11月1日生效，系个人信息保护领域的基础性法律。《个人信息保护法》以法律的形式，确立了处理个人信息的五大基本原则：合法正当必要诚信原则、目的限制原则、公开透明原则、质量原则、责任与安全原则等。《个人信息保护法》建立了个人信息的基本保护制度，以及敏感个人信息特别处理规则，个人信息跨境规则，规定了企业对个人信息进行分类管理、安全措施、制定应急预案，进行合规审计，个人信息保护影响评估，配合个人行使个人信息权益等义务和相应责任。

4.《中华人民共和国科学技术进步法》

2021年12月24日，全国人大常委会修订通过《科学技术进步法》（主席令第103号），于2022年1月1日施行，系科学技术领域的基础性法律。《科学技术进步法》建立了科技创新、开发、应用领域基本制度，落实重大改革创新举措，提供宽松的科研环境，保护自由探索等权益；鼓励科学技术研究开发，支撑实现碳达峰碳中和目标；深入推进科技体制改革，加强基础研究，优化区域创新布局，完善科技型企业上市融资制度；完善科技人员管理制度，鼓励科研单位激励科学技术人员等。

基于《网络安全法》《数据安全法》和《个人信息保护法》等法律法规对网信部门的授权，以及国务院《关于授权国家互联网信息办公室负责互联网信息内容管理工作的通知》（国发(2014)33号），国家网信办负责统筹协调全国网络安全工作、网络数据安全、个人信息保护工作以及全国互联网信息内容的监督管理和执法工作。

5.《互联网信息服务算法推荐管理规定》

2021年12月31日，国家网信办等四部门发布了《算法推荐规定》（网信办令第9号）。《算法推荐规定》于2022年3月1日起生效。《算法推荐规

定》明确了算法推荐服务的涵盖范围，及其监管机关；为服务提供者明确了信息服务规范，要求服务提供者应当坚持主流价值导向，积极传播正能量，并采取措施防范和抵制不良或违法信息；明确了用户权益保护要求，如公示服务的基本原理、目的意图和主要运行机制等，并保障算法选择权，禁止不合理的差别待遇；同时规定了推荐算法备案及其监管工作。

6.《互联网信息服务深度合成管理规定》

2022年11月25日，国家网信办联合工信部和公安部两部门共同颁布了（网信办第12号令，"**《深度合成规定》**"）。《深度合成规定》于2023年1月10日起生效。作为我国第一部针对深度合成服务治理的专门性部门规章，《深度合成规定》系AIGC领域的核心监管规定。《深度合成规定》明确了深度合成技术的范围、监管部门和基本原则，鼓励相关行业组织加强行业自律，履行消费者及其个人信息的保护义务，明确服务提供者和技术支持者的管理规范和主体责任。

据艾瑞咨询，2023年中国AIGC产业应用的主要分为两大板块，消费级终端和行业解决方案。其中，消费级终端包括文本、图片、音视频等内容生成，以及行业解决方案包括游戏、媒体/影视、广告营销、电商、金融、在线教育、金融等方面。也就是说，目前的生成式AI服务，主要以通过互联网向上网用户提供信息的服务活动，如文字、图片、音视频或其他文化产品的方式，常以网页版、移动应用程序（包括SDK和API）等方式直接向用户提供服务，或者利用互联网用户账号和AIGC在网络平台面向社会公众发布文字、图片、音视频等信息内容。因此，我国目前与AIGC直接相关的法律法规、政策和标准还包括：

序号	法律	制定部门	生效日期
1	《中华人民共和国反电信网络诈骗法》	全国人大常委会	2022/12/1
2	《中华人民共和国著作权法》	全国人大常委会	2021/6/1
3	《中华人民共和国未成年人保护法》	全国人大常委会	2021/6/1
4	《中华人民共和国刑法》及其修正案	全国人民代表大会	2021/3/1
5	《中华人民共和国民法典》	全国人民代表大会	2021/1/1
6	《中华人民共和国反恐怖主义法》	全国人大常委会	2018/4/27
7	《中华人民共和国广告法》	全国人大常委会	2015/9/1
8	《中华人民共和国国家安全法》	全国人大常委会	2015/7/1
9	《中华人民共和国治安管理处罚法》	全国人大常委会	2013/1/1
10	《关于维护互联网安全的决定》	全国人大常委会	2009/8/27
序号	**行政法规**	**制定部门**	**生效日期**
11	《中华人民共和国著作权法实施条例》	国务院	2013/3/1
12	《互联网信息服务管理办法》	国务院	2011/1/8
13	《计算机信息网络国际联网安全保护管理办法》	国务院	2011/1/8
序号	**国务院规范性文件**	**制定部门**	**生效日期**
14	《中共中央　国务院关于构建数据基础制度更好发挥数据要素作用的意见》	中共中央，国务院	2022/12/2

序号	部门规章	制定部门	生效日期
15	《网站平台受理处置涉企网络侵权信息举报工作规范》	中共中央网络安全和信息化委员会办公室（**"中央网信办"**）	2023/8/10
16	《互联网弹窗信息推送服务管理规定》	国家网信办等	2022/9/30
17	《关于支持建设新一代人工智能示范应用场景的通知》	科学技术部	2022/8/12
18	《移动互联网应用程序信息服务管理规定》	国家网信办	2022/8/1
19	《互联网用户公众账号信息服务管理规定》	国家网信办	2021/2/22
20	《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》	国家网信办等	2020/11/30
21	《国家新一代人工智能标准体系建设指南》	国家标准化管理委员会等	2020/7/27
22	《网络信息内容生态治理规定》	国家网信办	2020/3/1
23	《网络音视频信息服务管理规定》	国家网信办等	2020/1/1
24	《互联网文化管理暂行规定》	原文化部	2017/12/15
25	《互联网新闻信息服务管理规定》	国家网信办	2017/6/1
26	《互联网视听节目服务管理规定》	原国家新闻出版广电总局	2015/8/28
27	《互联网用户账号名称管理规定》	国家网信办	2015/3/1

序号	部门规范性文件	制定部门	生效日期
28	《网络安全标准实践指南——生成式人工智能服务内容标识方法》	全国信息安全标准化技术委员会秘书处（**“信安标委秘书处”**）	2023/8/25
29	《关于加强互联网信息服务算法综合治理的指导意见》	国家网信办等	2021/9/17
30	《关于进一步压实网站平台信息内容管理主体责任的意见》	国家网信办等	2021/9/15

另外，我们注意到《国务院2023年度立法工作计划》预备将人工智能法草案提请全国人大常委会审议。

/PART 002

值得AIGC服务提供者关注的合规建议

综合上述规定，我们建议服务提供者在提供AIGC前或过程中，履行以下几方面的义务：

（1）建立健全对AIGC的内容审核机制，提高AIGC的准确性和可靠性，建立违法和不良信息识别模型并进行模型优化训练等。具体来说，尤其应当关注《网络安全法》《互联网信息服务管理办法》等法律，并避免出现该等规定明确列举的以下信息：(一)反对宪法所确定的基本原则的；(二)危害国家安全，泄露国家秘密，颠覆国家政权，推翻社会主义制度，煽动分裂国家，破坏国家统一的；(三)损害国家荣誉和利益的；(四)宣扬恐怖主义、极端主义，宣扬民族仇恨、民族歧视，破坏民族团结的；(五)破坏国家宗教政策，宣扬邪教

和封建迷信的；(六)散布谣言，扰乱社会秩序，破坏社会稳定的；(七)散布淫秽、色情、赌博、暴力、凶杀、恐怖或者教唆犯罪的；(八)编造、传播虚假信息扰乱经济秩序和社会秩序；(九)侮辱或者诽谤他人，侵害他人名誉、隐私、个人信息、知识产权和其他合法权益的；(十)含有法律、行政法规禁止的其他内容的。

（2）按照《深度合成规定》等法律法规，一方面完成使用者的事前真实身份信息认证；另一方面在《深度合成规定》指定的智能对话、合成人声、人脸/姿态操控或替换、沉浸式拟真等场景中，参考《生成式人工智能服务内容标识方法》对生成文本、图片、音频、视频等内容时对生成内容进行显式或隐式标识。

（3）在发现违法内容或违法使用者时，应当采取包括向有关部门报告等应对措施。具体来说，如发现违法内容，服务提供者还应当及时采取停止生成、停止传输、消除等处置措施，并采取模型优化训练等措施进行整改等；如发现使用者从事违法活动的，服务提供者还应当依法依约采取警示、限制功能、暂停或者终止向其提供服务等处置措施，并保存有关记录。我们认为，相较于《AI服务办法》征求意见稿中要求的"在3个月内通过模型优化训练等方式防止再次生成"，《AI服务办法》提供了更为可行的"采取模型优化训练"等整改措施要求。

（4）采取有效措施反对不良信息的传播，如建立健全辟谣机制、设置便捷的用户申诉和公众投诉、举报入口等，以确保提升生成式AI服务的透明度，提高生成内容的准确性和可靠性。

另外，如果AIGC涉及网络信息、网络音视频信息的制作与传播方面的内容，或者内容涉及具有舆论属性或社会动员能力的、新闻信息服务、文化服务、视听节目等，服务提供者还应当重点关注上述国家网信办《网络信息内容生态治理规定》《网络音视频信息服务管理规定》《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》等规定。

/PART 003

社会公德与伦理道德体系梳理

《AI服务办法》所提到的尊重"社会公德和伦理道德"，其范围较为模糊。我们简要梳理了以下几部文件，试图探索一个社会公德和伦理道德的指引和可落地的抓手，供读者参考。

序号	部门规范性文件	制定部门	生效日期
1	《中国关于加强人工智能伦理治理的立场文件》	外交部	2022/11/17
2	《关于加强科技伦理治理的意见》	中共中央办公厅、国务院办公厅	2022/3/20
3	《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（**"AI委员会"**）	2021/9/25
4	《网络安全标准实践指南——人工智能伦理安全风险防范指引》	全国信息安全标准化技术委员会秘书处（**"信安标委秘书处"**）	2021/1/5
5	《新一代人工智能治理原则——发展负责任的人工智能》	科技部AI委员会	2019/6/17
6	《新一代人工智能发展规划》	国务院	2017/7/8

序号	其他文件	发布部门	发布日期
7	《科技伦理审查办法（试行）》	科学技术部等	2023/9/7 (2023/12/1 生效)
8	《人工智能法示范法（专家建议稿）》	中国社会科学院	2023/8/15
9	《人工智能伦理治理标准化指南》	国家人工智能标准化总体组等	2023/3/16

经过上述《新一代人工智能发展规划》《新一代人工智能治理原则——发展负责任的人工智能》等阶段的探索，结合国内外对AI技术及伦理规范的实践经验和教训，科技部在《国家新一代人工智能标准体系建设指南》中指出，AI安全/伦理标准应当贯穿于包括AI基础共性、支撑技术与产品、基础软硬件平台、关键通用技术、关键领域技术、产品与服务和行业应用等在内的AI标准体系结构其他部分，为AI建立合规体系。

2021年1月，信安标委秘书处在《网络安全标准实践指南——人工智能伦理安全风险防范指引》将AI伦理安全风险总结为以下五大方面：（1）失控性风险，如AI的行为与影响超出服务提供者预设、理解和可控的范围，对社会价值等产生负面影响；（2）社会性风险：不合理使用AI而对社会价值等方面产生负面影响；（3）侵权性风险：AI对人的基本权利，包括人身、隐私、侵权性风险财产等造成侵害或产生负面影响；（4）歧视性风险：AI对人类特定群体具有主观或客观偏见，影响公平公正、造成权利侵害或负面影响；（5）责任性风险：AI相关各方行为失当、责任界定不清，对社会信任、社会价值等方面产生负面影响。

2021年9月，AI委员会发布的《新一代人工智能伦理规范》提出了“增进

人类福祉、促进公平公正、保护隐私安全、确保可控可信、强化责任担当、提升伦理素养"等6项AI基本伦理规范；形式上，伦理规范包括管理规范、研发规范、供应规范和使用规范等。

2022年3月，中共中央办公厅、国务院办公厅印发《关于加强科技伦理治理的意见》，提出"科技伦理是开展科学研究、技术开发等科技活动需要遵循的价值理念和行为规范，是促进科技事业健康发展的重要保障"，并明确了以下五大类科技伦理原则：增进人类福祉、尊重生命权利、坚持公平公正、合理控制风险和保持公开透明。《人工智能伦理治理标准化指南》对前述《意见》进行进一步细化。

2023年4月，科技部等印发《科技伦理审查办法（试行）》（征求意见稿，正式版本已于2023年10月8日发布，12月1日起施行，要求从事AI科技活动的企业设立科技伦理（审查）委员会，并向科技伦理主管部门提交年度工作报告。8月，社科院发布《人工智能法示范法1.0（专家建议稿）》（2023年9月7日更新至1.1版），规定了以人为本、安全、公开透明可解释、可问责、公平平等、绿色原则和促进发展创新等七项原则。在管理制度方面，该示范法还提出分类管理制度和负面清单管理制度，即社科院建议国家AI主管机关制定负面清单，要求开展清单内AI研发、提供活动应当获得相应行政许可。需要注意的是，《人工智能法示范法》虽然处于专家建议稿阶段，暂不具有法律效力，但是仍然值得作为AIGC伦理问题的向导而引起关注。

/PART 004
值得AIGC服务提供者关注的伦理建议

结合上述文件等各项规定，我们推荐相关企业就AIGC关注以下几个方面的伦理准则：

（1）符合我国主流社会价值观。具体来说，服务提供者可以基于提升社会价值的目标，明确AIGC的内容边界，预估其可能的影响范围，并采取措施为必要的AIGC确认社会主义核心价值观的解释依据；

（2）尊重并保护个人合法权益。具体来说，服务提供者一方面可以拒绝或避免开发以损害他人权益为主要目的的，或者容易受到恶意利用的AIGC技术；另一方面，在为生成式AI添加环境感知并独立生成AIGC相关功能时，或者检测到AIGC涉及公共服务、健康卫生、金融服务等涉及不特定多数人的利益时，服务提供者可以对用户进行特别提示，并采取必要措施保护个人人身、隐私、财产等权利，如当检测到AIGC可能影响个人权利时，相应内容应当具有清晰、明确、可查的法律法规等依据，同时采取措施避免泄露个人信息、个人隐私等可能侵害个人权益的内容。

（3）积极推动AIGC伦理安全建设，如建立健全覆盖管理、研发、供应、使用等全生命周期的风险治理体系、事件应对体系等。具体来说，针对面向用户的AIGC，服务提供者可以采取建立验证算法、风险预警、记录和回溯机制等必要措施，持续监测和降低风险；同时定期分析风险监控报告并反馈和优化管理机制，完善治理体系；另一方面，在企业内部持续进行防范措施宣传培训工作。服务提供者可以建立事件应对体系，设立人工紧急干预机制、中止应用机制、救济金基金等必要保障机制；明确事故处理流程，确保可以在AI伦理安全风险发生时作出及时响应，如停止问题产品生产、召回问题产品等；如条件允许，可以设置企业内救济金基金，或通过购买保险服务等必要手段，对引起的损失提供救济。

（4）在提升AI系统、产品和服务的可解释性、可控性和透明度的同时，可以通过用户协议、站内信、弹窗甚至公告等形式及时、准确、完整、清晰、无歧义地向用户说明AI系统、产品或服务的功能、局限、安全风险和可能的影响，以便用户准确判断AIGC是否存在误导、偏见等情况；并建议用户以良好目的使用AI、充分体现AIGC的积极作用，不应以有损社会价值、个人权利等

目的恶意使用。另一方面，企业应以清楚明确且便于操作的方式向用户提供拒绝、干预及停止使用AI相关系统、产品或服务的机制；在用户拒绝或停止使用后，应尽可能为用户提供非人工智能的替代选择方案；同时向用户提供清楚明确且便于操作的投诉、质疑与反馈机制，并提供包含人工服务在内的响应机制，进行处理和必要补偿。

参考文献：

1.最高人民法院司法案例研究院，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》全文及官方解读，https://mp.weixin.qq.com/s/kRgz7eaZdvolPovO7cF0rw，最后访问于2023年9月3日。

2.艾瑞，2023年中国AIGC产业全景报告，艾瑞咨询，https://mp.weixin.qq.com/s/xz8HbD34CFr70mbxnv1lmw，最后访问于2023年9月3日。

3.陈伟、颜炳琳、葛正一，《生成式人工智能服务管理暂行办法》正式施行，框定我国AIGC监管思路，https://www.junhe.com/legal-updates/2243，最后访问于2023年9月3日。

4.普华永道，政策解码：从合规领域视角探寻AIGC服务监管的创新与治理边界，普华永道，https://mp.weixin.qq.com/s/eFIsiL1NEIw16S0jyyp6EQ，最后访问于2023年9月3日。

5.胡珉琦：人工智能伦理：有原则不等于能治理，中国科学报，https://news.sciencenet.cn/sbhtmlnews/2022/6/369642.shtm，最后访问于2023年9月3日。

6.张平：人工智能伦理反思：风险与应对，中国社会科学报，https://epaper.csstoday.net/epaper/read.do?m=i&iid=6311&eid=44129&sid=204130&idate=12_2022-05-31，最后访问于2023年9月3日。

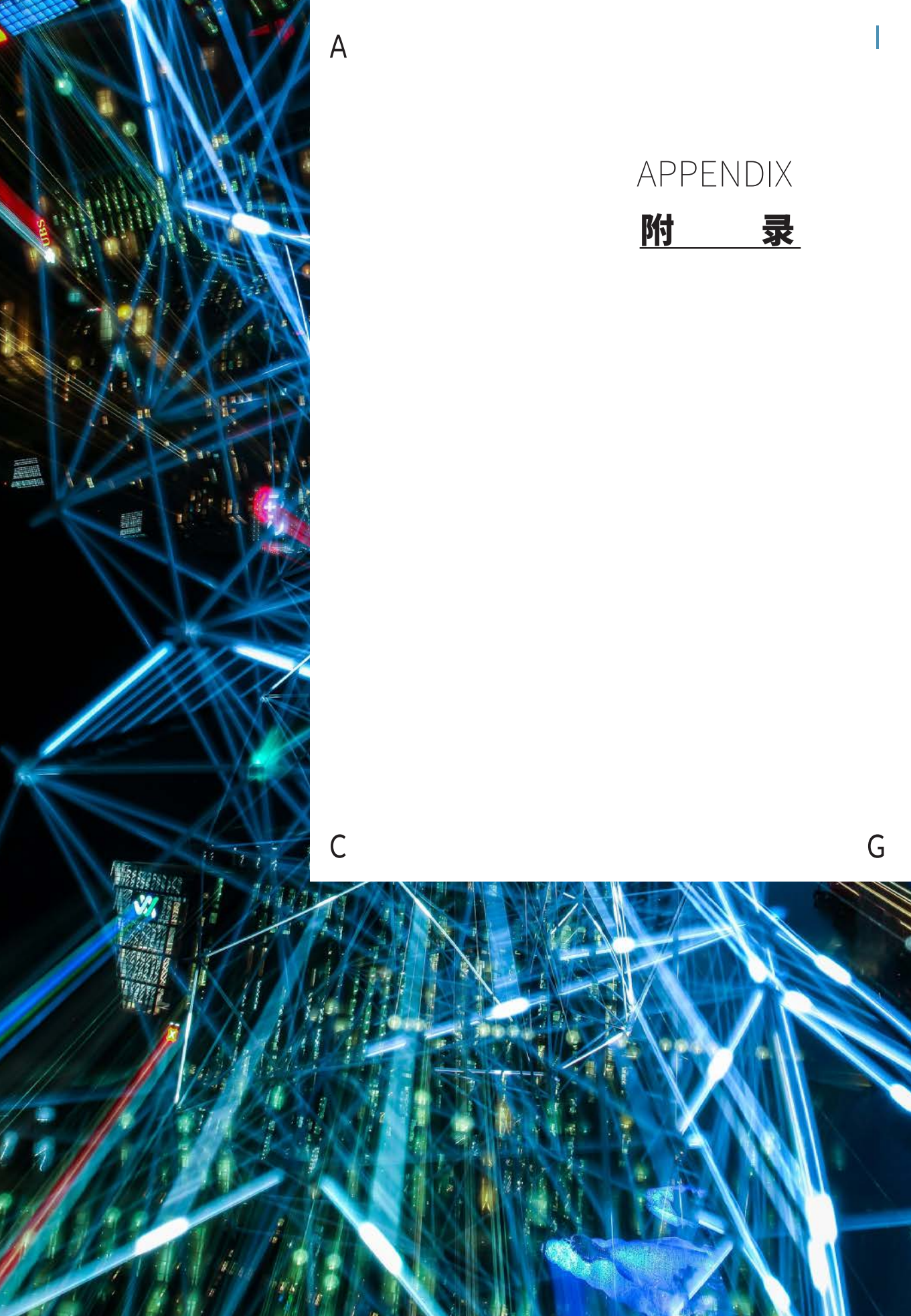
7.刘永谋等，人工智能的伦理挑战与科学应对，光明日报，https://news.gmw.cn/2023-04/10/content_36486021.htm，最后访问于2023年9月3日。

8.池骋，人工智能治理 | 世界卫生组织《人工智能伦理与治理指南》评述，清华大学智能法治研究院，https://mp.weixin.qq.com/s/UEQiIESCpH-VXO_EgQoe3ZQ，最后访问于2023年9月3日。

9.董子涵、任凤琴，困境与超越：人工智能的伦理审视，科技智囊，https://mp.weixin.qq.com/s/l6qzhrr6rs9uzzkyh0H4_g，最后访问于2023年9月3日。

10.韩冰意、苏中，“信息茧房”、隐私外泄，如何应对人工智能带来的伦理风险？，阿里研究院，https://mp.weixin.qq.com/s/Ok2f4SKKx-Sc6Nehr4-aBEg，最后访问于2023年9月3日。

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



A

C

G

APPENDIX

附　　录

生成式人工智能（“AIGC”）合规清单

陈际红 陈煜烺

数据合规			
检查要点	**合规指引**	**合规成果**	**相关法律法规**
数据质量	**1.在模型训练阶段，AIGC服务提供者应采取有效措施保证训练数据质量，增强训练数据的真实性、准确性、客观性、多样性。** - 增强数据集的技术设计，提升数据集的真实性、准确性、客观性、多样性等； - 制定数据质量标准与预处理操作指引等制度文本和操作规程。	- 数据质量标准 - 训练数据预处理操作指引	《生成式人工智能服务管理暂行办法》第7条
数据源合法性	**2.在模型训练阶段，针对数据爬取活动，AIGC服务提供者应识别可能面临的法律风险，制定数据爬取合规手册（含合规检查清单及操作指引）。** - **根据自身数据爬取活动的手段、内容及后续使用行为，制作合规检查清单，识别潜在法律风险：** **1)爬取手段：** （1）刑事风险：为越过被爬取网站的反爬虫技术措施而采取反反爬虫技术且获取数据，可能构成"非法获取计算机信息系统数据罪"； （2）如果因爬取行为造成被爬网站服务器不能正常运行，可能构成"破坏计算机信息系统罪"； （3）不正当竞争：如果违反目标网站Robots协议，或通过IP代理、伪造UA（User-Agent）等形式，或其他反反爬虫措施实施爬取，扰乱被爬网站的正常运营，则可能构成"妨碍、破坏其他经营者的网络产品或者服务正常运行"的不正当竞争行为，亦可能会有进一步的刑事责任风险；	- 数据爬取合规手册 - 数据源合法性审查规范 - 数据处理协议/条款 - 数据源合法性尽调报告 - 数据合作方定期审计报告	《生成式人工智能服务管理暂行办法》第7条 《个人信息保护法》《著作权法》《反不正当竞争法》《刑法》等

检查要点	合规指引	合规成果	
	（4）著作权保护：如果未经权利人许可，故意避开或者破坏权利人设定的技术措施，以爬取他人享有著作权的作品，可能承担民事侵权责任。 **2)爬取内容：** （1）个人信息保护：如果爬取内容构成个人信息，需满足透明性要求并具备合法性基础，避免爬取未合法公开的个人信息，否则可能构成"违法收集个人信息"的行为； （2）著作权保护：如果未经权利人许可，爬取他人享有著作权的作品，则此等爬取作品至自身数据库的行为可能构成对复制权的侵犯； （3）商业秘密：如果爬取内容由于保密技术措施存在而具有秘密性，则可能构成商业秘密，违规爬取可能构成"侵犯商业秘密"的行为。 **3)爬取数据的后续使用：** （1）个人信息保护：如果爬取已公开的个人信息并用于后续模型训练，则需确保满足法律关于处理已公开个人信息的要求，例如，在合理范围内处理且在对个人权益有重大影响时取得其同意； （2）著作权保护：如果未经权利人许可，爬取他人享有著作权的作品，进行分析、处理、演绎，并在后续输出成果中展示，亦可能侵犯他人著作权； （3）不正当竞争：如果爬取数据后的模型训练、提供AIGC服务行为被视为"对被爬网站构成实质性替代"，或者"爬取'衍生数据'构成'搭便车'"，则可能受到《反不正当竞争法》的规制。		

检查要点	合规指引	合规成果	相关法律法规
	- **制作供业务人员使用的数据爬取操作指引，指引中可明确要求业务人员：** （1）爬取公开的、非保密的前台数据，不应爬取非公开的后台数据； （2）避免采用破解密码、伪造设备IP绕过服务器身份校验、IP代理、伪造UA等技术手段以绕过或破解网站采取的保护数据的技术措施； （3）控制爬取的频率、流量等，不得因爬取数据导致被爬网站无法正常响应，甚至瘫痪。 **3.在模型训练阶段，AIGC服务提供者应建立数据源及数据模型审查机制，降低违法风险。** - 审查数据提供方是否具备提供数据的法定或议定依据，如涉及模型合作，服务提供者需审查模型是否具备合法来源。针对个人信息，应审查其是否具备提供个人信息的合法性基础；针对非个人信息，应确保所提供的数据不存在权属瑕疵； - 与数据提供方签署数据处理协议或在业务合同中加入数据保护条款； - 针对数据提供方开展合规尽调，就其数据保护层面的安全能力、组织管理、操作规程等进行调查； - 要求数据提供方配合定期开展合规审计等。		

检查要点	合规指引	合规成果	相关法律法规
个人信息保护	**4.AIGC服务提供者应梳理各阶段所涉及的个人信息处理活动，并重点关注以下主要个人信息合规要点：** - 履行透明性及合法性要求：应告知服务使用者个人信息处理规则，并取得相应合法性基础（通常为同意或履行合同所必需）；如涉及未满十四周岁未成年人用户，应制定专门个人信息处理规则； - 最小必要原则和数据保留要求：不得非法留存能识别服务使用者身份的信息，包括服务使用者输入信息及使用记录，保存期限应为实现处理目的所需的最短时间； - 个人信息共享：制定第三方数据合规管理规范；共享个人信息前应具备合法性基础，并与其签订数据处理协议，明确各方数据处理关系（共同处理、委托处理、对外提供等）及责任分配； - 开展个人信息保护影响评估； - 制定个人信息处理者行权响应规范，建立个人信息主体行权机制，保障服务使用者行使个人信息主体权利； - 如涉及个人信息出境，见下文。	- 合法性基础审查规范 - 隐私政策或其他告知文本 - 未满十四周岁未成年人的个人信息处理规则 - 数据处理协议/条款 - 第三方数据合规管理规范 - 个人信息共享记录 - 个人信息保护影响评估报告 - 个人信息主体行权响应规范	《生成式人工智能服务管理暂行办法》第11条 《个人信息保护法》第7、13、14、15、17、19、20、21、22、23、31、44、50、55、59条

检查要点	合规指引	合规成果	相关法律法规
数据出境	**5.AIGC服务提供者应识别数据出境场景，遵循出境合规机制、履行告知义务并取得合法性基础。** - 识别数据出境场景：例如，是否涉及境外服务提供者直接面向境内提供服务；或是否涉及由境内服务提供者通过API等形式集成境外AIGC技术向境内提供服务； - 遵循出境合规机制：根据主体类型、出境数据类型及量级，开展数据出境安全评估/个人信息出境标准合同备案/个人信息保护认证； - 履行告知义务(如涉及个人信息)：除前述"个人信息保护"的告知要求外，针对个人信息出境，还需告知境外接收方的名称或者姓名、联系方式、处理目的、处理方式、个人信息的种类以及个人向境外接收方行权的方式和程序等事项； - 具备合法性基础（如涉及个人信息）：除前述"个人信息保护"部分中关于合法性基础的要求外；针对个人信息出境，应取得单独同意（如以同意作为合法性基础）； - 开展个人信息保护影响评估（如涉及个人信息）。	- 数据出境场景识别规范 - 数据出境风险自评估报告（如适用） - 个人信息出境标准合同（如适用） - 个人信息保护影响评估报告 - 数据跨境传输协议	《生成式人工智能服务管理暂行办法》第20条 《数据安全法》第31条 《网络安全法》第37条 《个人信息保护法》第13、38、39、55条

检查要点	合规指引	合规成果	相关法律法规
数据泄露	**6.根据AIGC产品数据处理活动的不同情形，AIGC服务提供者和服务使用者应制定不同的防范数据泄露风险的措施。** **- 在模型训练、应用运行、模型优化等各阶段：** 针对AIGC自身产品/系统可能发生数据安全事件，根据相关规定制定数据安全应急预案，在发生数据安全事件时及时履行上报及通知义务，并对第三方系统/产品供应商采取安全层面的合规管理措施； **- 在应用运行阶段：** 1)针对服务提供者，为避免含有商业秘密或其他敏感数据（个人信息、重要数据等）的数据进入AIGC学习资料库，并进而在对外输出内容过程中发生数据泄露，一方面可通过用户协议或其他文本提示服务使用者应避免输入公司商业秘密、个人信息或重要数据等敏感数据，另一方面应采取一定技术手段防止此等因输出内容而导致的数据泄露事件发生； 2)针对服务使用者，应识别高风险应用场景并开展使用AIGC技术的必要性审查（例如，是否可使用其他替代性技术），并采取针对性措施；通过员工手册等内部制度要求员工不得输入保密数据等手段避免数据泄露。	- 数据安全应急预案 - 数据安全事件处置记录/报告 - 用户协议或其他文本 - 风险识别及必要性审查规范 - 员工手册	《生成式人工智能服务管理暂行办法》第20条 《数据安全法》第29条 《网络安全法》第25条

知识产权保护			
检查要点	**合规指引**	**合规成果**	**相关法律法规**
模型训练侵权风险	**7.AIGC服务提供者应避免使用未经授权的作品开展模型训练，并建立"标注＋退出机制""合理使用"或其他免责事由的论证规范以及侵权响应机制。** - 使用受《著作权法》保护的作品进行模型训练前，应尽可能取得版权方授权； - 建立"标注＋退出机制"。考虑到前置授权许可的成本及现实难度，可在AIGC生成内容对涉及作品的使用情况作出标注和说明，并允许作者"选择退出"，以增加作者对于其作品使用的感知和控制，缓释风险； - 建立"合理使用"或其他免责事由的论证规范及侵权响应机制。考虑到前置授权许可的成本及现实难度，AIGC服务提供者应在内部建立"合理使用"或其他免责事由的论证规范，并针对外部版权方建立侵权响应机制，以缓释风险。	- 著作权授权文本 - 标注＋退出机制 - "合理使用"或其他免责事由的论证规范 - 侵权响应机制	《生成式人工智能服务管理暂行办法》第7条 《著作权法》第24条 《著作权法实施条例》第21条
开源风险	**8.如涉及基于开源模型开发自身AIGC产品，或将自研AIGC模型以开源形式对外提供，应关注开源风险。** - 如基于开源模型开发自身AIGC产品，应关注不同的开源许可类型，尽可能选择限制性条款较少的开源模型，重点关注开源条款的"传染性"，并遵守开源协议约定，避免侵权或违约风险； - 如基于自研AIGC模型以开源形式对外提供，应构建企业内部开源管理体系，制定开源许可协议，并采取相关技术措施，避免模型的底层技术（know-how）发生泄露或被反向工程。	- 企业开源管理体系 - 开源许可协议	/

检查要点	合规指引	合规成果	相关法律法规
输入内容及输出内容的权利安排及限制	**9.AIGC服务提供者应与服务使用者通过用户协议等文本，明确约定输入内容及生成内容的权利安排及限制，避免潜在争议，约定内容包括但不限于：** - 输入内容的权利安排，例如服务提供者是否可对输入内容作后续的衍生使用（例如用于模型优化）； - 生成内容的权利安排，例如服务使用者是否享有生成内容的归属（包括知识产权），以及服务提供者是否对生成内容享有使用权（例如用于模型优化）； - 生成内容的使用限制，例如服务使用者不得将其用于从事违法活动或实施侵权行为。	- 用户协议等文本	《生成式人工智能服务管理暂行办法》第7条
反不正当竞争与反垄断	**10.AIGC服务提供者和使用者应建立AIGC正当使用规范，不得利用算法、数据、平台等优势，实施垄断和不正当竞争行为。**	- AIGC正当使用规范	《生成式人工智能服务管理暂行办法》第4条

监管合规			
检查要点	**合规指引**	**合规成果**	**相关法律法规**
主体责任	**11.AIGC服务提供者应当落实信息安全主体责任，建立相关管理制度及投诉、举报机制。** - 建立用户注册、算法机制机理审核、科技伦理审查、信息发布审核、数据安全、个人信息保护、反电信网络诈骗、应急处置等管理制度； - 建立辟谣机制，发现利用AIGC服务制作、复制、发布、传播虚假信息的，应当及时采取辟谣措施，保存有关记录，并向网信部门和有关主管部门报告； - 建立投诉、举报机制，设置便捷的投诉、举报入口，公布处理流程和反馈时限，及时受理、处理投诉举报并反馈处理结果。	- 信息安全系列管理制度 - 辟谣机制 - 投诉、举报机制	《生成式人工智能服务管理暂行办法》第15条 《深度合成管理规定》第7、11、12条
用户管理	**12.AIGC服务提供者可通过产品功能设计和产品页面的文本展示，进行服务使用者行为管理。** - 注册管理：参照“后台实名、前台自愿”的原则落实真实身份信息认证制度； - 使用管理：与服务使用者签订服务协议，明确权利义务；公开产品使用规则（公开适用的人群、场合、用途），采取有效措施防范未成年人用户过度依赖或者沉迷AIGC服务； - 违法行为管理：发现服务使用者利用AIGC服务从事违法活动的，应当依法依约采取警示、限制功能、暂停或者终止向其提供服务等处置措施，保存有关记录，并向有关主管部门报告。	- 服务使用者管理机制 - 用户协议 - 产品使用规则 - 防沉迷措施	《生成式人工智能服务管理暂行办法》第9、10、14条 《互联网信息服务深度合成管理规定》第9条 《互联网用户账号名称管理规定》第5条 《互联网用户账号信息管理规定》第9条

检查要点	合规指引	合规成果	相关法律法规
内容安全	**13.AIGC服务提供者应履行网络信息内容生产者责任，建立内容审核机制及处置机制。** - 履行网络信息内容生产者责任，不得生成违法信息，并应采取措施防范和抵制生成不良信息；建立技术或人工内容审核机制，对输入内容和输出内容进行审查； - 建立违法和不良信息特征库，完善入库标准、规则和程序，记录并留存相关网络日志； - 如发现违法信息和不良信息，应当及时采取停止生成、停止传输、消除等处置措施，采取模型优化训练等措施进行整改，并向网信部门和有关主管部门报告；对服务使用者依法依约采取警示、限制功能、暂停服务、关闭账号等处置措施。	- 内容审核机制 - 违法和不良信息特征库 - 违法信息和不良信息处置机制	《生成式人工智能服务管理暂行办法》第9、14条 《互联网信息服务深度合成管理规定》第10条 《网络信息内容生态治理规定》第4、6、7条
数据标注	**14.AIGC服务提供者应建立数据标注制度。** - 制定清晰、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，提升遵法守法意识并监督其规范开展标注工作； - 如涉及标注服务外包，应要求外包公司协助履行相关法律要求。	- 数据标注规则 - 数据标注人员操作规范 - 数据标注培训	《生成式人工智能服务管理暂行办法》第8条

检查要点	合规指引	合规成果	相关法律法规
透明度要求	**15.AIGC服务提供者应采取措施提升AIGC服务的透明度，公示算法原理、意图、运行机制等。** - 以产品/服务说明等方式，公示AIGC算法的基本原理、目的意图和主要运行机制等； - 采取技术措施，提高数据标注、清洗及模型训练的透明性、可解释性，并以产品/服务说明等方式公示。	- 产品/服务说明 - 提高透明度的技术措施	《生成式人工智能服务管理暂行办法》第4条 《互联网信息服务算法推荐管理规定》第16条
标识要求	**16.AIGC服务提供者应对生成内容进行显著标识，避免公众混淆或误认。** - 参照《网络安全标准实践指南 生成式人工智能服务内容标识要求（征求意见稿）》等制定内部标识规则，并依法进行标识，例如：在显示区域下方或使用者输入信息区域下方持续显示提示文字，或在显示区域的背景均匀添加包含提示文字的显式水印标识；由人工智能生成图片、视频时，应采用在画面中添加提示文字的方式进行标识，提示文字宜处于画面的四角，所占面积应不低于画面的0.3%或文字高度不低于20像素，提示文字应至少包含"由人工智能生成"或"由 AI 生成"等信息等。	- 生成内容标识规则	《生成式人工智能服务管理暂行办法》第12条 《互联网信息服务深度合成管理规定》第16、17条 《网络安全标准实践指南——生成式人工智能服务内容标识要求》

检查要点	合规指引	合规成果	相关法律法规
算法备案	**17.具有舆论属性或社会动员能力的AIGC服务提供者应依法开展算法备案，公示备案信息并履行备案变更、注销义务（如涉及）。** - 应在提供服务之日起10个工作日内通过互联网信息服务算法备案系统提交算法备案申请。备案内容包括服务提供者的名称、服务形式、应用领域、算法类型、算法自评估报告、拟公示内容等； - 在对外提供服务的网站、应用程序等的显著位置标明备案编号并提供公示信息链接； - 如涉及备案信息变更或终止服务等情形，应依法办理备案变更、注销手续。	- 算法安全自评估报告	《生成式人工智能服务管理暂行办法》第17条 《互联网信息服务算法推荐管理规定》第24、25、26条
安全评估	**18.具有舆论属性或社会动员能力的AIGC服务提供者上线AIGC服务前，应依法自行开展安全评估，并向网信部门和公安机关提交安全评估报告。** - 上线AIGC服务前，应依法自行开展安全评估；评估内容参见《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》第7条； - 应将安全评估报告通过全国互联网安全管理服务平台提交所在地地市级以上网信部门和公安机关。	- 安全评估报告	《生成式人工智能服务管理暂行办法》第17条 《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》第3、5、7条

检查要点	合规指引	合规成果	相关法律法规
双新评估	**19.AIGC服务提供者上线AIGC服务前，应开展"互联网新技术新业务安全评估"（又称"双新评估"）。** - 上线AIGC服务前，参照《YD/T 3169-2020 互联网新技术新业务安全评估指南》《YD/T 3738-2020 互联网新技术新业务安全评估实施要求》等规定，开展双新评估。	- "双新评估"报告	《YD/T 3169-2020 互联网新技术新业务安全评估指南》 《YD/T 3738-2020 互联网新技术新业务安全评估实施要求》
其他行政许可/备案	**20.AIGC服务提供者应根据具体业务模式，依法取得相应行政许可/备案，包括但不限于：** - 公安联网备案； - 互联网信息服务许可/备案（ICP许可/备案）；以及其他增值电信许可，例如EDI证、SP证等； - 其他特殊行政许可，例如《网络文化经营许可证》《网络出版服务许可证》《信息网络传播视听节目许可证》等。	/	《生成式人工智能服务管理暂行办法》第23条 《中华人民共和国电信条例》第7条 《电信业务分类目录（2015年版）》（2019年修订） 《计算机信息网络国际联网安全保护管理办法》第12条 《互联网信息服务管理办法》 《互联网文化管理暂行规定》 《网络出版服务管理规定》 《互联网视听节目服务管理规定》

检查要点	合规指引	合规成果	相关法律法规
外资准入限制	**21.如涉及外商投资AIGC服务，应根据所处国家或地区（例如是否位于香港/澳门）、具体细分行业领域要求，并结合相关业务模式应取得的行政许可/备案（同前），遵循适用的外资准入要求。**	/	《生成式人工智能服务管理暂行办法》第23条 《外商投资电信企业管理规定》 《外商投资准入特别管理措施（负面清单）》 《自由贸易试验区外商投资准入特别管理措施 （负面清单）》 《内地与香港/澳门关于建立更紧密经贸关系的安排》
反歧视	**22.AIGC服务提供者应采取有效措施防止歧视。** - 建立多维度的数据集并在算法模型中引进公平性等考量，以防止歧视； - 此等反歧视措施应涵盖算法设计、训练数据选择、模型生成和优化等过程。	- 反歧视的技术措施	《生成式人工智能服务管理暂行办法》第4条

检查要点	合规指引	合规成果	相关法律法规
科技伦理审查	**23.AIGC服务提供者应关注科技伦理审查的有关流程和审查要点，尽快部署科技伦理审查制度。** - 根据《科技伦理审查办法（试行）》及相关规定的要求，尽快开展合规部署。	- 科技伦理审查规范	《科学技术进步法》 《生成式人工智能服务管理暂行办法》第1、4条 《算法推荐管理规定》第7条 《深度合成管理规定》第7条 《科技伦理审查办法（试行）》 《关于加强科技伦理治理的意见》

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

总编

陈际红

龚乐凡

张炯

编委（按姓氏笔画排序）

王飞

王红燕

吴小旭

张鹏

周洋

赵刚

顾萍

斯响俊

蔡荣伟

蔡鹏

樊晓娟

中伦研究院出品

WWW.ZHONGLUN.COM