生成式人工智能运营合规全景问答（一）：范围与概览

作者：傅鹏、刘嘉纯、赵卿梦、俞沁、钱学悦

2023 年 4 月 11 日发布的《生成式人工智能服务管理办法（征求意见稿）》（“**《生成式人工智能服务办法》**”）针对近期引起高度关注的 ChatGPT、Midjourney 等模式的生成类人工智能应用，提出了一系列合规要求，引起业界的广泛讨论。而《生成式人工智能服务办法》的规管路径，在一些方面也沿袭甚至直接引用了《互联网信息服务算法推荐管理规定》（“**《算法推荐管理规定》**”）和《互联网信息服务深度合成管理规定》（“**《深度合成管理规定》**”）的监管框架。

面对人工智能的“iPhone 时刻”和走向通用人工智能的“关键一步”，面对许多商业主体在大型语言模型的基础层、中间层和应用层都有机会开辟自身具有特色的服务和产品的商业背景，甚至面对有的人士提出的“绝大部分传统应用的服务模式都值得根据大型语言模型重做一遍”的产业前瞻，熟悉生成式人工智能运营合规的要点，在牌照备案、数据安全、个人信息保护、运营合规、商业合同要点和知识产权保护等方面进行全景式的了解，是生成式人工智能开发者和服务提供者需要特别关注的内容。

1. 当我们讨论《生成式人工智能服务办法》及其监管时，其实是在讨论哪些受到监管的技术与活动，这类技术的实践应用主要有哪些？

生成式人工智能，往往体现为利用人工智能技术生成或合成一定的成果。在《生成式人工智能服务办法》中，将服务形态和规制的活动描述为“研发、利用生成式人工智能产品，面向中华人民共和国境内公众提供服务的”，而其中的“生成式人工智能”则指“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”。

以上范围比较宽泛，和此前颁布的涉及生成合成类技术的规定也略有不同。例如，《算法推荐管理规定》监管的活动是“在中华人民共和国境内应用算法推荐技术

提供互联网信息服务”。而“应用算法推荐技术”指“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”。

在《深度合成管理规定》中，具体要求“在中华人民共和国境内应用深度合成技术提供互联网信息服务，适用本规定”，而“深度合成技术”是指利用深度学习、虚拟现实等生成式算法制作文本、图像、音频、视频、虚拟场景等网络信息的技术，包括但不限于：（一）篇章生成、文本风格转换、问答对话等生成或者编辑文本内容的技术；（二）文本转语音、语音转换、语音属性编辑等生成或者编辑语音内容的技术；（三）音乐生成、场景声编辑等生成或者编辑非语音内容的技术；（四）人脸生成、人脸替换、人物属性编辑、人脸操控、姿态操控等生成或者编辑图像、视频内容中生物特征的技术；（五）图像生成、图像增强、图像修复等生成或者编辑图像、视频内容中非生物特征的技术；（六）三维重建、数字仿真等生成或者编辑数字人物、虚拟场景的技术。”

相比《算法推荐管理规定》和《深度合成管理规定》而言，《生成式人工智能服务办法》希望管理和规制的活动，更为宽泛。该文件不继续局限于利用“算法”生成结果的活动，而是一般性地规定为“基于算法、模型、规则”生成结果的活动。这一规定变化，一定程度上反映了实践中遇到的难点与困惑。实际项目中，在很多场景下，对于何谓“算法”，对于什么活动真正属于使用了“算法”，并没有很明确的定义与区分。所以，在《生成式人工智能服务办法》颁布之前，许多项目实践中，也倾向于认为，如果使用了一定的相对复杂的规则、方法，生成了服务结果，则应当视为受到“算法”规则规制的活动。

从上述法规文本规定的边界和形态出发，可以基本明确，典型的文本生成应用（例如 ChatGPT），图像生成应用（例如 Midjourney），音频生成应用（例如 DeepMusic），视频生成应用（例如 Deepfake）等，都在《生成式人工智能服务办法》等规定意在规管的“生成式人工智能”范围之内。

2. 除了 ChatGPT、Midjourney 等典型的基于大型语言模型或者预训练方式研发的生成类产品外，《生成式人工智能服务办法》是否也监管 AR/VR 等场景融合、合成和增强类产品？

在文本逻辑上，AR/VR 等典型的虚拟场景类、现实增强类服务，是否被含括在《生成式人工智能服务办法》所述的“生成式人工智能”范围内，存在较为模糊的理解空间。

《生成式人工智能服务办法》规定的“生成式人工智能”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”，一定程度上包含了 VR 应用的部分场景中运用到的图片、声音、视频等生成物。另外，上述规定采取了开放型的列举模式，使用“等内容”进行兜底，在文意上并不排斥将 AR/VR 场景应用含括在内。

然而，从《生成式人工智能服务办法》全文意在监管的活动来看，并没有明显迹象显示这一文件意在监管传统意义上以现实增强和虚拟场景生成作为主要功能的 AR/VR 服务。例如，该办法第七条和第十七条多次提到产品的预训练和优化训练数据，再如办法第十九条规定服务提供者对从事网络炒作、恶意发帖跟评、制造垃圾邮件、编写恶意软件，实施不正当的商业营销等行为，应当暂停或者终止服务。上述规定针对的还是典型的内容结果生成式人工智能服务，甚至意在指向基于大型语言模型、以预训练方式形成产品基础的人工智能服务模型。上述规定及其意图，在 AR/VR 服务场景中并不容易找到典型的适用维度。

因此，取决于后续主管机关颁布的正式生效的《生成式人工智能服务办法》规定，其宜被理解为指向基于大型语言模型、以预训练方式形成产品基础的人工智能服务，而并不直接针对典型的 AR/VR 服务。

当然，AR/VR 服务中，也完全可能嵌入生成式服务内容（例如 VR 游戏中的 NPC 对话），对于这些典型的生成式服务内容，毫无疑问需要受到《生成式人工智能服务办法》的规管。此外，AR/VR 服务，视具体形态，也较有可能需要遵守《算法推荐管理规定》和《深度合成管理规定》的要求，履行这两个规定中的合规义务。

当然，本次国家互联网信息办公室（“**网信办**”）对《生成式人工智能服务办法》的发文风格偏向于敏捷式监管，意在以相对垂直的切口，规制监管相对热点的话题，并以较快速度释放征求意见稿甚至是最终生效的规定条文，并可能在相对短的时间内针对监管重点的衍化与发展，进行相对快速的“迭代式”修订。如果遵循这样的做法，也不排除后续的“类 AR/VR”服务被明确纳入生成式人工智能服务的可能，如果条文与实践朝这一方向发展，则业界也需要相应跟进与关注。

3. 在哪里开展生成式人工智能服务，需要受到《生成式人工智能服务办法》的监管？

《生成式人工智能服务办法》规定，研发、利用生成式人工智能产品，面向中华人民共和国境内公众提供服务的，适用该办法。所以，只要是面向中国境内公众提供服务，就需要遵守《生成式人工智能服务办法》的规定。

因此，如果生成式人工智能服务的提供者位于境外，但是面向中国境内公众提供服务，例如面向中国境内公众进行宣传，设置有中文页面，开通了面向和便利中国境内公众进行付款的支付通道，都可能被认定为面向中国境内公众提供服务。这一方面呼应（或反证）了实践中的一些具体做法。例如，OpenAI 在账户注册环节，要求用户输入手机号，而输入中国大陆地区手机号的用户不被允许注册，一定程度上可以理解为 OpenAI 并不意在面向中国大陆的公众提供服务。在一些其他项目中，有的服务提供者不允许来自中国大陆 IP 地址的用户进行注册，也可以理解为并不意在面向中国大陆公众提供服务。

另一方面，上述规定的方式也与《中华人民共和国个人信息保护法》（“**《个保法》**”）和《网络数据安全管理条例（征求意见稿）》（“**《网数条例》**”）的规制路径相类似。《个保法》第 3 条规定，在中国境外处理中国境内自然人个人信息的活动，如果以向境内自然人提供产品或者服务为目的，或者分析评估境内自然人的行为，则境外处理者的活动也适用《个保法》的监管。《网数条例》第 2 条规定，在中国境外处理中国境内个人和组织数据的活动，如果以向境内提供产品或者服务为目的，或者分析评估境内个人、组织的行为，则需要遵守《网数条例》的规定。

所以，《生成式人工智能服务办法》的监管逻辑相比于此前颁布的《算法推荐管理规定》和《深度合成管理规定》都更加宽泛。例如，《算法推荐管理规定》要求，在中国境内应用算法推荐技术提供互联网信息服务，需要适用该规定的监管。《深度合成管理规定》也规定，在中国境内应用深度合成技术提供互联网信息服务，需要受到该规定的监管。上述两个规定都要求在中国境内提供服务才受到相应监管。相比之下，《生成式人工智能服务办法》并不要求这一地域限定，扩大了自身监管的对象范围。

4. 生成式人工智能服务，总体而言需要注意哪些方面的法律合规要求或法律问题？

生成式人工智能服务需特别关注如下几个主要方面的问题。如下从整体方面进行概括介绍，我们将在后续的几篇分析中，进行具体详尽的介绍与解析。

A. 取得必要的资质许可和完成必要的备案、登记

- **算法备案**

在近几个月生成式人工智能应用爆火之前，已经存在不少生成合成类服务，主管部门将其纳入算法监管的范畴，发布了《算法推荐管理规定》进行管理。其中规定，具有舆论属性或者社会动员能力的算法推荐服务提供者应当在提供服务之日起十个工作日内通过互联网信息服务算法备案系统填报服务提供者的名称、服务形式、应用领域、算法类型、算法自评估报告、拟公示内容等信息，履行备案手续。《生成式人工智能服务办法》也规定，利用生成式人工智能产品向公众提供服务，应当按照《算法推荐管理规定》履行算法备案和变更、注销备案手续。

- **安全评估**

《生成式人工智能服务办法》规定，利用生成式人工智能产品向公众提供服务前，应当按照《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》向国

家网信部门申报安全评估。此外，生成式人工智能服务也属于《深度合成管理规定》规管的活动，而《深度合成管理规定》也规定，深度合成服务提供者开发上线具有舆论属性或者社会动员能力的新产品、新应用、新功能的，应当按照国家有关规定开展安全评估。因此，生成式人工智能服务提供者应当根据上述规定进行安全评估，并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的主管部门。

- **增值电信业务经营许可**

生成式人工智能服务，根据具体模式，需要判断是否涉及增值电信服务的形态。如果不涉及增值电信服务形态，则不需要取得增值电信业务经营许可证；如果涉及，则需要取得。

一般而言，如果服务提供者仅进行软件、算法开发，授权或部署给客户使用，而不涉及其他更具互动性（特别是允许用户进行互动或发送信息的功能）的线上服务形态，则可能不需要取得增值电信业务经营许可证。

如果服务提供者在提供生成式服务功能时，允许用户发布信息、进行互动（典型如聊天室、论坛、即时通信等功能），则可能需要取得 B25 类（信息服务业务）增值电信业务经营许可证（即俗称的“ICP 证”）。如果服务提供者以生成式服务作为引流手段，同时经营电商平台，则可能需要取得 B21 类（在线数据处理与交易处理业务，特别是其中的“交易处理业务”）增值电信业务经营许可证。服务提供者的其他一些服务和业务形态，如果属于某种增值电信业务活动，则需要取得其他类型的相应子类别牌照。

- **文化视听类许可**

生成式人工智能服务如果仅仅生成问答文字，或者针对特定任务进行文字回答，一般不属于需要取得《网络出版服务许可证》《网络文化经营许可证》《信息网络传播视听节目许可证》《广播电视节目制作经营许可证》等文化视听类许可牌照的活动。

如果服务提供者通过网络向用户提供网络出版物，则需要根据《网络出版服务管理规定》等规则的要求，取得《网络出版服务许可证》。对于什么是“网络出版物”，《网络出版服务管理规定》规定了比较宽泛的范围，但是在实践中，值得把握两个特点：（1）《网络出版服务管理规定》也明确规定，网络出版物是“具有编辑、制作、加工等出版特征的数字化作品”，所以需要具有“出版特征”；（2）实践认定过程中，对于通过对话进行的文字输出，一般篇幅相对较短（当然也有例外），也缺乏编辑、加工等特征，一般不认定为出版物。

如果服务提供者提供互联网文化产品及其服务，则需要取得《网络文化经营许可证》。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（例如网络音乐娱乐、网络演出剧（节）目、网络表演、网络艺术品、网络动漫等）。

如果服务提供者提供了互联网视听节目服务（例如制作、编辑、集成并通过互联网向公众提供视音频节目，以及为他人提供上载传播视听节目服务），则需要取得《信息网络传播视听节目许可证》。

如果服务提供者从事广播电视节目制作经营活动，需要取得《广播电视节目制作经营许可证》。

B. 在业务运营过程中遵守合规要求

综合《生成式人工智能服务办法》《深度合成管理规定》及其他规定的要求，生成式人工智能服务提供者在业务运营过程中，需要关注如下专项合规要求：

- 用户注册与管理：需落实用户实名制注册；注意内容角度的用户防沉迷；合理引导用户使用生成式人工智能生成内容，不利用生成内容损害他人合法权益，不进行商业炒作、不正当营销。

- 数据安全、个人信息保护：需关注训练数据的合法性；保护用户的注册个人信息和使用过程中输入的个人信息；对用户的输入信息和使用记录进行保护。

除非法律法规另有规定，不留存推断用户身份的输入信息，不根据用户输入信息和使用情况进行画像，不向他人提供用户输入信息。此外，由于生成式人工智能是典型的数据驱动型业务，在训练、使用过程中也往往涉及大量个人信息，甚至是比较敏感的个人信息，所以需要全面遵守《个保法》和《中华人民共和国数据安全法》《中华人民共和国网络安全法》的网络安全、数据安全及个人信息保护要求。

- 内容生态治理：服务提供者需要注意避免提供带有歧视性的内容生成；产品研制中采用人工标注时，应当制定符合要求的标注规则，对标注人员进行必要培训，抽样核验标注内容的正确性；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。

- 其他：需要采取算法机制机理审核，信息发布审核，平台内容管理，建立健全辟谣机制，并对人工智能生成内容进行不影响用户使用的标识。

C. 关注知识产权风险，保护知识产权

生成式人工智能服务提供者需要关注，在训练阶段，使用的训练资料以及生成的结果不应侵犯他人的知识产权。其中，对于生成的结果，视服务的形态和使用的技术模型的不同，侵权风险可能相差较大。在相当一部分场景下，特别是在文字生成任务中，基于大型语言模型的生成式人工智能服务产生的结果往往不是训练语料的原文复述，所以具有相对较低的知识产权侵权风险。但是，在一些图片生成模型的生成效果中，可能出现与训练图像较为近似的生成图像结果。生成式人工智能服务提供者需要在模型设置、训练资料的选取、以及训练和调优过程中注意此类生成结果侵权的风险。

尽管相当一部分生成式人工智能模型的基础逻辑原理都已经披露在公开的论文当中，但是在生成式人工智能模型的具体开发过程中，以及将生成式人工智能模型封装为产品以及与其他软件结合提供服务的过程中，可能涉及到开源代码的使用。生成式人工智能服务提供者需要注意开源代码许可证的许可条款要求，特别

需要关注开源代码许可证的许可条款是否要求“开源传染性”，即要求使用了开源代码的开发结果的代码必须开源，以及其他附随要求。

此外，人工智能生成的结果是否具有著作权，由谁享有著作权，已经成为学术界和实务界各方人士争论颇多的一个问题。不论主张生成结果没有著作权，还是由生成服务使用者享有著作权（因为使用者在提供指令以及多次调整指令过程中，可能对生成结果的表达形成了一定程度的“独创性”），还是由生成式人工智能的开发者（服务提供者）享有著作权（因为生成结果毕竟是人工智能写就的产物），都面临较大争议和挑战。这一令人困惑和保守争议的问题尽管不直接关系到生成式人工智能服务提供者是否合法合规地提供了服务，但是也与其用户（生成式服务的使用者）能不能在更多维度和更多方面使用生成结果并主张自己的一定权利具有很大的关系。

以上，是我们在《生成式人工智能服务管理办法（征求意见稿）》为首的监管规定背景下，对生成式人工智能服务运营合规在范围和合规要求方面的概括介绍。后续，我们将从“牌照与备案”“业务运营合规”“商业协议关注点”“知识产权关注点”等角度，分别对运营过程中的合规要求和重点考虑要素进行更加详细的介绍。