《生成式人工智能服务管理暂行办法》的五大要点解读

实务问题	个人信息保护（网络安全与数据合规->个人信息保护），网络信息内容审查（网络安全与数据合规->网络信息内容审查）		
发文日期	2023-09-13	作者	陈晓霞
作者来源	北京大成律师事务所		
法学分类	高新技术（科技法->高新技术）		
行业分类	科学与技术		

正文内容:

生成式人工智能技术在技术领域被称为AIGC（AI Generated Content），其主要通过大规模的训练数据，辅之以功能完善的编码器和解码器对其进行编码学习与解码生成，再通过有效的评估机制后生成用户所需的内容。例如，先前全球流行的ChatGPT（Chat Generative Pre-trained Transformer）即为典型的生成式人工智能，其可以与用户进行普通聊天、完成信息咨询、撰写诗词文章等。生成式人工智能具有易于使用、提高效率、节约成本等显著优势，可以辅助企业与个人做出更精准的决策，但与此同时，也面临着诸多挑战，数据偏差、监管伦理、隐私安全等已为人工智能领域进一步发展亟需解决的问题。

2023年7月10日，国家互联网信息办公室同国家发展和改革委员会等五部委、国家广播电视总局共同发布了《生成式人工智能服务管理暂行办法》（以下简称"暂行办法"），并于2023年8月15日正式实施。此前，国家互联网信息办公室曾于2023年4月11日发布了《生成式人工智能服务管理办法（征求意见稿）》（以下简称"征求意见稿"），短短三个月时间，从征求意见稿到正式稿的出台，体现了国家对于人工智能技术飞速发展的关切，也是我国立法在生成式人工智能这一新兴领域的勇敢尝试。暂行办法具有五大要点，本文将逐一解读。

一、适用范围明确化

暂行办法第二条从正反两面对生成式人工智能的适用范围进行了明确，规定"利用生成式人工智能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，适用本办法。"同时，还明确排除了不适用的场景，一类是"新闻出版、影视制作、文艺创作等领域"，另一类是"行业组织、企业、教育和科研机构、公共文化机构、有关专业机构等研发、应用生成式人工智能技术，未向境内公众提供生成式人工智能服务的情形"。

首先，暂行办法将适用范围限定在"提供服务"层面，豁免了暂行办法对于研发阶段的规制，减缓了企业研发模型及其相关技术的合规要求，旨在鼓励生成式人工智能在多领域多渠道的进一步探索。其次，对于"境内公众"如何理解？"公众"仅指不特定的多数人，还是亦包含企业。一种观点认为，企业不应囊括在公众的范畴中，另一种观点认为，不特定的多数企业也应被理解为"公众"。

我们认为，根据暂行办法包容审慎又谋求发展突破的制定目的而言，应将其理解为不包括向特定企业提供服务的行为，仅指不特定多数人可以普遍接触到的生成式人工智能模型，如此，有利于技术的进一步发展。但该情形究竟是否落入暂行办法的规制范围，仍有待实践中进一步明确。

二、监管政策二元化

暂行办法对生成式人工智能服务的监管政策进行了更明确的规定。主要提出两种监管政策，一是根据生成式人工智能服务的风险高低进行分类分级监管，二是根据生成式人工智能服务适用的不同领域进行行业部门监管。这两种监管政策相辅相成，共同促进立法对人工智能的体系化监管进一步加强。

1. 分类分级监管

暂行办法在第三条中明确提出了分类分级监管："国家坚持发展和安全并重、促进创新和依法治理相结合的原则，采取有效措施鼓励生成式人工智能创新发展，对生成式人工智能服务实行包容审慎和分类分级监管。"同时在十六条中再次对分类分级监管的原则进行了确认。所谓分类分级监管是指，根据生成式人工智能应用的服务场景，对应用中可能出现的风险及影响从高到低进行排序，再根据不同风险等级划定不同的监管方式。我国在《中华人民共和国数据安全法》（以下简称"《数据安全法》"）中也采取了此种监管思路，在《数据安全法》的二十一条中规定："国家建立数据分类分级保护制度，根据数据在经济社会发展中的重要程度，以及一旦遭到篡改、破坏、泄露或者非法获取、非法利用，对国家安全、公共利益或者个人、组织合法权益造成的危害程度，对数据实行分类分级保护。"

虽然我国目前并未提出进一步的分类分级依据以及操作办法，但《国务院2023年度立法工作计划》已将《人工智能法草案》列为预备提请全国人大常委会审议的法律草案。我们有理由期待，在《人工智能法》出台后，关于分类分级监管的原则会得到进一步的发展与落实。

2. 行业部门监管

除了分类分级监管的原则外，暂行办法在第十六条中明确规定"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。"该条规定意味着对于生成式人工智能的监管，从"有"开始转向"精"，立法摒弃了大包大揽式的笼统规定，转而赋予各行业根据自身行业特点制定更为贴合行业特征的监管规定的权利，便于将生成式人工智能的监管落实到各行业的实践中。

行业部门监管的政策，使得人工智能监管法规具有更强的针对性，是对于人工智能强技术性特征的回应与迎合，便于各行业根据自身需求制定更为合理的生成式人工智能服务监管政策。例如，在公安领域中应用的生成式人工智能，如何促使其在追踪犯罪中打击范围更全面、如何保障犯罪追踪的过程中不侵犯他人的隐私安全，即为公安部门在制定监管政策时需要重点考量的要素。

三、外资准入严格化

暂行办法第二十条、二十三条对生成式人工智能服务的境外提供者进行了较大的限制。其中，第二十条规定："对来源于中华人民共和国境外向境内提供生成式人工智能服务不符合法律、行政法规和本办法规定的，国家网信部门应当通知有关机构采取技术措施和其他必要措施予以处置。"境外的生成式人工智能服务提供者在对境内提供服务时，会受到较大的限制。但值得思考的是，除了明确的对境外服务提供者的限制之外，对于境内服务提供者采用境外生成式人工智能模型提供服务的情形，法律并未禁止，暂行办法亦未将其排除出适用范围。

暂行办法第二十三条对行政许可与外商投资进行了进一步规定："法律、行政法规规定提供生成式人工智能服务应当取得相关行政许可的，提供者应当依法取得许可。外商投资生成式人工智能服务，应当符合外商投资相关法律、行政法规的规定。"一方面，现行法律并未对提供生成式人工智能服务本身是否需要取得行政许可进行规定，若从法无禁止即自由的角度而言，普通领域适用生成式人工智能服务并不需要取得特别的许可。但若涉及到金融安全、医药行业、互联网文化经营等特殊领域，该领域的准入本身即需要取得有关部门的行政许可，那在该领域适用生成式人工智能服务可能随之面临许可类的监管要求。另一方面，根据现行外商投资准入规定，生成式人工智能并非外商禁止投资或限制投资的项目，但考虑到人工智能技术的日新月异，此种空白可能并非是默认的许可，而是立法滞后性的体现。根据暂行办法第二十条的条文意旨，对境外服务提供者进行限制的本质是在人工智能领域加强对外资的控制，那么类推而言，由外商投资的生成式人工智能服务也应受到限制。

四、法律责任具体化

暂行办法相较征求意见稿，对法律责任的相关内容进行了进一步的具体规定。在扩大责任主体的同时，对于不同主体需承担的不用义务及责任也进行了进一步规定，在促进责任主体多元化的同时增强了责任内容的多样化，为暂行办法的进一步实施提供了良好的基础。

暂行办法从十六条到二十一条，分别对不同主体在生成式人工智能服务中的法律责任进行了详细规定。第十六条着眼于网信、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，规定各部门均有职责对生成式人工智能进行管理与规制；第十七条对生成式人工智能服务的提供者的责任进行了明确，规定其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；第十八条规定了使用者的责任，当使用

者发现生成式人工智能服务不符合法律、行政法规和暂行办法规定的，有权向有关主管部门投诉、举报；第十九条在明确相关部门应当对生成式人工智能服务的提供者开展监督检查的同时，明确了提供者在面对监督检查时的责任，其应当依法予以配合，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。此外，该条还要求参与监督检查的机构和人员对其在检查过程中获知的国家秘密、商业秘密、个人隐私等内容具有保密的义务；第二十条对境外的生成式人工智能服务提供者进行了规制，若其不符合法律、法规及暂行办法的规定向境内提供服务的，国家网信部门有权对其进行查处；第二十一条延续了《征求意见稿》的思路，通过转致条款，将违反暂行办法的处罚措施转为依据《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》《中华人民共和国科学技术进步法》等法律、行政法规的规定予以处罚，体现了法律规范之间的衔接性，同时促进了法律责任的进一步精细化。

五、合规要求全面化

暂行办法从原则性的角度对生成式人工智能服务提出了要求，并在内容中进一步规定了提供者的合规要求和使用者的合规要求。其中，暂行办法明显更侧重于对生成式人工智能服务的提供者予以规制，从研发到应用的诸多过程中都确立了需满足的合规要求。

1. 原则性要求

暂行办法第四条对提供和使用生成式人工智能服务从三个角度规定了原则性的合规要求：

其一，从国家社会角度，第四条第一款规定"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"，这是从最基本也最宏观的角度对生成式人工智能服务进行规制；

其二，从主体平等角度，第四条的第二款规定在算法设计、训练数据选择、模型生成和优化、提供服务的过程中需采取有效措施防止产生民族、信仰、国别、地域、性别、年龄、职业、健康等歧视，第三款从反垄断和反不正当竞争的角度提出了要求；

其三，从隐私安全角度，第四条的第四款和第五款分别从肖像权、名誉权、荣誉权、隐私权和个人信息权益的保护和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这两方面对生成式人工智能的服务提供者和使用者做出了规制。

2. 提供者的合规要求

暂行办法对生成式人工智能的服务提供者规定了较为全面的规范要求，体现了立法在谋求人工智能技术发展的同时，对于人工智能在应用合规的重视。

第一，暂行办法第七条规定了生成式人工智能的服务提供者在模型训练与优化阶段的合规义务，包括提供者应该使用具有合法来源的数据和基础模型、不得侵害他人依法享有的知识产权、涉及个人信息时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及增强训练数据的真实性、准确性、客观性、多样性等要求。

第二，暂行办法第八条对研发过程中，服务提供者的数据标注义务进行了详细规定，要求提供者应当制定符合要求的清晰、具体、可操作的标注规则、开展数据标注质量评估、抽样核验标注内容的准确性等，建议企业合法合规开展数据标注及质量评估工作，并注意工作留痕，也需关注监管方是否会进一步发布相关标准。

第三，暂行办法第九条规定了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，在涉及个人信息时，应当依法承担个人信息处理者责任，履行个人信息保护义务。同时，该条还规定提供者应当与生成式人工智能服务的使用者签订服务协议，明确双方权利义务。该项规定将签订协议的实践惯例上升到法定义务层面，增强了生成式人工智能服务在应用过程中的合规要求。

第四，暂行办法第十条规定了提供者关于未成年人的合规义务。其要求提供者对生成式人工智能服务的用途场景进行明确，

并采取有效措施防范未成年人用户过度依赖或者沉迷生成式人工智能服务。该条款体现了立法对于未成年人的保护，同时也彰显了对于提供者在生成式人工智能服务技术实际应用过程中，对使用者进行管理的合规义务的加强。

第五，暂行办法第十一条规定了提供者关于个人信息保护的合规义务。其规定服务提供者应当对使用者的输入信息和使用记录依法履行保护义务，不得收集非必要个人信息，不得非法留存能够识别使用者身份的输入信息和使用记录，不得非法向他人提供使用者的输入信息和使用记录。同时要求提供者应当依法及时受理和处理个人关于查阅、复制、更正、补充、删除其个人信息等的请求。个人信息保护问题一直是立法的重中之重，此次暂行办法的出台从法律层面认定了生成式人工智能服务的提供者可能成为个人信息处理者，从而对其赋予了必要的个人信息保护义务，在回应实践关切的同时，也对提供者的合规义务进行了进一步要求。

第六，暂行办法第十二条和第十三条分别规定了提供者的标识义务和网络安全义务。第十二条规定提供者应当按照《互联网信息服务深度合成管理规定》对图片、视频等生成内容进行标识。第十三条规定提供者应当在其服务过程中，提供安全、稳定、持续的服务，保障用户正常使用。

第七，暂行办法第十四条规定了提供者针对违法行为进行整改的合规义务。其规定当提供者发现违法内容时，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。若发现使用者利用生成式人工智能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。

第八，暂行办法第十五条规定了提供者应建立与公众互通渠道的合规义务。其要求提供者应当建立健全投诉、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、处理公众投诉举报并反馈处理结果，旨在通过面向公众的投诉举报机制的建立，维护生成式人工智能服务的合法合规性。

第九，暂行办法第十七条规定了提供者有关评估备案的合规义务。对于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，显著加强了在人工智能具有较强舆论导向的情况下，提供者应完成的合规要求。

3. 使用者的合规要求

相较于生成式人工智能服务的提供者，立法对于使用者的合规要求大大降低。除了暂行办法第四条对于提供者和使用者均适用的原则性规定外，还在第十八条中赋予了使用者具有投诉举报的权利："使用者发现生成式人工智能服务不符合法律、行政法规和本办法规定的，有权向有关主管部门投诉、举报"。该法条并非严格意义上对于使用者的合规要求，而是以生成式人工智能服务的合规体系出发，从使用者的角度促使生成式人工智能服务的合法合规。

《生成式人工智能服务管理暂行办法》的出台，是我国立法在人工智能领域的一大进步，在包容审慎的原则基础上，限缩了适用范围的规定，对监管机制进行体系化设计，具体规定了不同主体的法律责任要求，并进一步明确了服务提供者和使用者的合规义务，促进了生成式人工智能服务合法合规性的完善。飞速发展的人工智能产业对法律规定的建立健全提出了紧迫的需求，法律规定的完善又进一步推进了人工智能产业的合法合规，二者相互促进、共同提升。在不久的将来，随着《人工智能法草案》的面世，相信人工智能领域的设计与监管将更加完善，将进一步推动具有中国特色的人工智能体系的建立。

本文作者

陈晓霞
大成北京　高级合伙人
xiaoxia.chen@dentons.cn
专业领域：公司与并购、争议解决、银行与金融

来源简介

北京大成律师事务所
官方网站:http://dacheng.com
北京大成律师事务所成立于1992年，是中国成立最早、规模最大的合伙制律师事务所之一。秉承“志存高远，海纳百川，跬步千里，共铸大成”的文化核心理念以及“全球资源、本土智慧”的开放式发展战略，大成始终致力于为国内外客户提供专业、全面、及时、优质、高效的法律服务。
经过三十余年的发展，大成已经成为在中国境内拥有48家办公室，服务范围覆盖国内全部省、自治区和直辖市，以及与世界范围内80余个国家及地区设有160多个办公室的全球最大律师事务所Dentons拥有优先合作关系的规范化、规模化、专业化、品牌化、国际化的大型综合性律师事务所。